The architecture of an empirical genotype-phenotype map

Aguilar-Rodríguez, José ; Peel, Leto ; Stella, Massimo ; Wagner, Andreas ; Payne, Joshua L

Abstract: Recent advances in high-throughput technologies are bringing the study of empirical genotype-phenotype (GP) maps to the fore. Here, we use data from protein-binding microarrays to study an empirical GP map of transcription factor (TF) -binding preferences. In this map, each genotype is a DNA sequence. The phenotype of this DNA sequence is its ability to bind one or more TFs. We study this GP map using genotype networks, in which nodes represent genotypes with the same phenotype, and edges connect nodes if their genotypes differ by a single small mutation. We describe the structure and arrangement of genotype networks within the space of all possible binding sites for 525 TFs from three eukaryotic species encompassing three kingdoms of life (animal, plant, and fungi). We thus provide a high-resolution depiction of the architecture of an empirical GP map. Among a number of findings, we show that these genotype networks are "small-world" and assortative, and that they ubiquitously overlap and interface with one another. We also use polymorphism data from Arabidopsis thaliana to show how genotype network structure influences the evolution of TF-binding sites in vivo. We discuss our findings in the context of regulatory evolution.

DOI: https://doi.org/10.1111/evo.13487

The architecture of an empirical genotype-phenotype map

Running title: The architecture of an empirical GP map

José Aguilar-Rodríguez[1,2,†], Leto Peel[3,4], Massimo Stella[5], Andreas Wagner[1,2,6,*], Joshua L. Payne[2,7,*]

1 Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

2 Swiss Institute of Bioinformatics, Lausanne, Switzerland

3 Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

4 Namur Center for Complex Systems, University of Namur, Namur, Belgium

5 Institute for Complex Systems Simulation, Department of Electronics and Computer Science, University of Southampton, UK

6 The Santa Fe Institute, Santa Fe, New Mexico, USA

7 Institute for Integrative Biology, ETH, Zurich, Switzerland

† Current affiliation: Department of Biology, Stanford University, Stanford, CA, USA; Department of Chemical and Systems Biology, Stanford University, Stanford, CA, USA

∗ Correspondence to: andreas.wagner@ieu.uzh.ch, joshua.payne@env.ethz.ch

Author contributions

J.A-R., A.W., and J.L.P designed research; J.A-R., L.P., M.S., and J.L.P. performed research; J.A-R., L.P., M.S., A.W., and J.L.P. analyzed data; J.A-R., A.W., and J.L.P. wrote the manuscript.

Acknowledgments

This work was partially carried out during the Complex Systems Summer School 2014 at the Santa Fe Institute, NM, USA. We thank Evandro Ferrada, Cole Mathis, and Renske Vroomans for exciting discussions and valuable input. J.A.-R. acknowledges support through the Forschungskredit program of the University of the Zurich (grant FK-14-076). L.P. acknowledges support through the F.R.S-FNRS and ARC "Mining and Optimization of Big Data Models" of the Federation Wallonia-Brussels. M.S. was supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1). A.W. acknowledges support by ERC Advanced Grant 739874, by Swiss National Science Foundation grant 31003A_172887, as well as by the University Priority Research Program in Evolutionary Biology at the University of

Zurich. J.L.P. acknowledges support through the Forschungskredit program of the University of the Zurich (grant K-74301-04-01), as well as Swiss National Science Foundation grants PZ00P3_154773 and PP00P3_170604.

Archiving statement

All data are freely available in a Dryad repository (doi:10.5061/dryad.5fb633t)

Abstract

Recent advances in high-throughput technologies are bringing the study of empirical genotype-phenotype (GP) maps to the fore. Here, we use data from protein binding microarrays to study an empirical GP map of transcription factor (TF) binding preferences. In this map, each genotype is a DNA sequence. The phenotype of this DNA sequence is its ability to bind one or more TFs. We study this GP map using genotype networks, in which nodes represent genotypes with the same phenotype, and edges connect nodes if their genotypes differ by a single small mutation. We describe the structure and arrangement of genotype networks within the space of all possible binding sites for 525 TFs from three eukaryotic species encompassing three kingdoms of life (animal, plant, and fungi). We thus provide a high-resolution depiction of the architecture of an empirical GP map. Among a number of findings, we show that these genotype networks are "small-world" and assortative, and that they ubiquitously overlap and interface with one another. We also use polymorphism data from *Arabidopsis thaliana* to show how genotype network structure influences the evolution of TF binding sites *in vivo*. We discuss our findings in the context of regulatory evolution.

Introduction

Evolution can be abstracted as an exploration of genotype space — the space of all possible genotypes (Maynard Smith 1970). This space is populated by intersecting sets of genotypes that each correspond to a distinct phenotype. The organization of genotype space into such genotype sets is described by the genotype-phenotype (GP) map (Burns 1970; Alberch 1991), an object of central importance in developmental and evolutionary biology, with important implications for medicine (Pigliucci 2010; Wagner and Zhang 2011; Lehner 2013).

Most of what we know about GP maps comes from computational models of biological systems (Lipman and Wilbur 1991; Schuster et al. 1994; Ciliberti et al. 2007; Rodrigues and Wagner 2009; Cotterell and Sharpe 2010; Salazar-Ciudad and Marín-Riera 2013; Greenbury et al. 2014). These include models that map RNA sequence genotypes onto secondary structure phenotypes (Schuster et al. 1994; Aguirre et al. 2011), simplified amino acid sequence genotypes onto lattice-based, structural phenotypes (Lipman and Wilbur 1991; Bornberg-Bauer and Chan 1999), regulatory circuit genotypes onto gene expression phenotypes (Ciliberti et al. 2007), and metabolic genotypes onto nutrient utilization phenotypes (Rodrigues and Wagner 2009). GP maps have also been studied in non-biological systems, including self-replicating computer programs (Fortuna et al. 2017), evolutionary algorithms (Hu et al. 2012), and field programmable gate arrays (Raman and Wagner 2011). Despite all that differentiates these systems, their GP maps have much in common (Ahnert 2017). First, they are many-to-one, meaning that multiple genotypes have the same phenotype. Second, the distribution of genotypes per phenotype is heavily skewed, such that most phenotypes are realized by few genotypes, and a few phenotypes are realized by many genotypes. Third, genotypes with the same phenotype tend to be mutationally interconnected, meaning that it is possible to transform any one of these genotypes into any other via a series of small mutations that preserve the phenotype. Such sets of mutationally interconnected genotypes are known as genotype networks (aka neutral networks (Schuster et al. 1994)). A fourth commonality is that the genotype networks of different phenotypes tend to overlap and interface with one another (Wagner 2008; Barve and Wagner 2013; Payne and Wagner 2013; Wagner 2014). We refer to the comprehensive description of the structure and arrangement of genotype networks within genotype space as the architecture of a GP map (Ferrada 2014).

The architecture of a GP map has important implications for evolution, influencing the rate of adap-

tation (Draghi et al. 2010; Manrubia and Cuesta 2015), the "findability" of genotypes and phenotypes in evolutionary searches (Cowperthwaite et al. 2008; McCandlish 2013; Schaper and Louis 2014), as well as their robustness and evolvability (Wagner 2008). It is therefore important to move beyond the study of GP maps derived from computational models, and to begin to study the architecture of GP maps that are derived from experimental data.

We currently know very little about the architecture of such empirical GP maps. The reason is that the genotype spaces of most biological systems are so large that it is not possible to experimentally assay a phenotype for all possible genotypes (Louis 2016). This is especially problematic when studying the architecture of a GP map, where it is necessary to assay a large number of phenotypes. Recent advances in high-throughput sequencing and chip-based technologies are beginning to mitigate this problem by providing localized descriptions of GP maps for macromolecules such as RNA and proteins (Rowe et al. 2010; Fowler et al. 2010; Hinkley et al. 2011; Jiménez et al. 2013; Melamed et al. 2013; Buenrostro et al. 2014; Findlay et al. 2014; Olson et al. 2014; Podgornaia and Laub 2015; Julien et al. 2016; Li et al. 2016; Puchta et al. 2016; Qiu et al. 2016; Sarkisyan et al. 2016). While insightful, these empirical GP maps still only describe a small subset of the genotype networks of a small number of phenotypes, and therefore cannot be used to characterize the architecture of a GP map.

In contrast, protein binding microarrays (Berger et al. 2006) provide comprehensive descriptions of transcription factor (TF) binding preferences to all possible, short DNA sequences (eight nucleotides in length), and such data are available for a large number of TFs (Weirauch et al. 2014). These data can therefore be used to describe the architecture of an empirical GP map at high resolution, in which each genotype is a DNA sequence (TF binding site), and the phenotype of this DNA sequence is its ability to bind one or more TFs. These are biologically important phenotypes, because TF binding is integral to the transcriptional regulation of gene expression, which underlies fundamental developmental, behavioral, and physiological processes in species as different as bacteria and humans (Ptashne and Gann 2002). What is more, DNA mutations that affect transcriptional regulation, including those in TF binding sites, may lead to evolutionary adaptations and innovations (Prudhomme et al. 2007; Wray 2007). Examples include binding site mutations that affect body plans in snakes (Guerreiro et al. 2013) and the discrimination of optical stimuli in fruit flies (Rister et al. 2015).

Characterizing the architecture of a GP map helps us to understand how such adaptations and innovations may arise. For example, genotype network structure provides information about how genetic

diversity accumulates in an evolving population (van Nimwegen et al. 1999). Combined with an understanding of how genotype networks interface with one another, this information provides insight into how mutations may bring forth new phenotypes. Similarly, by characterizing the overlap of genotype networks with one another, it is possible to study biological phenomena as different as exaptation (Barve and Wagner 2013), plasticity (West-Eberhard 2003), and multifunctionality (Payne and Wagner 2013). In the context of TF binding sites, such overlap is indicative of "crosstalk", a phenomenon in which multiple TFs compete for the same binding site, which may lead to incorrect gene activation or repression, as well as the titration of TFs away from their target sites (Friedlander et al. 2016).

In previous work, protein binding microarray data were used to characterize the topologies and topographies of genotype networks of TF binding sites (Payne and Wagner 2014; Aguilar-Rodríguez et al. 2017). The goals of these studies were to characterize the relationship between robustness and evolvability in TF binding sites (Payne and Wagner 2014), and to understand how mutation and natural selection might navigate such networks toward high-affinity binding sites (Aguilar-Rodríguez et al. 2017). To accomplish these goals, the genotype networks of TF binding sites were constructed and studied individually, providing localized characterizations of genotype space. Here, we extend this earlier work by providing a global and more detailed characterization of this genotype space for hundreds of TFs across three kingdoms of life, thus describing the architecture of an empirical GP map at high resolution. Moreover, we do so at two levels of granularity: that of individual TFs and of entire DNA binding domain structural classes.

Materials and Methods

In vitro data

We studied data from protein binding microarrays (Berger et al. 2006), a chip-based technology that measures the *in vitro* binding preferences of a TF to all possible 32,896 double-stranded DNA sequences of length eight. There are $\frac{4^8-4^4}{2} + 4^4 = 32,896$ such sequences, because each is merged with its reverse complement and because there are 4^4 sequences that are identical to their reverse complement and therefore cannot be merged. We refer to these DNA sequences as TF binding sites (or simply "sites") because we study the capacity of these sequences to bind TFs. The binding preferences of a TF are reported as a list of protein binding microarray enrichment scores (E-scores), one per binding site (Berger

et al. 2006). The E-score is a non-parametric, rank-based variant of the Wilcoxon-Mann-Whitney statistic that ranges from -0.5 to 0.5. It correlates with a TF's relative dissociation constant, and is therefore used as a proxy for relative binding affinity (Berger et al. 2006; Badis et al. 2009). We used this proxy to discriminate sites that specifically bind a TF via hydrogen bond donors and acceptors (E-score > 0.35), from unbound sites or sites that bind a TF non-specifically, for example via the TF's affinity for the DNA backbone (E-score ≤ 0.35). We chose the threshold $\tau = 0.35$, which has been used in previous studies (Nakagawa et al. 2013; Payne and Wagner 2014; Aguilar-Rodríguez et al. 2017), because it corresponds to a low false discovery rate (below 0.001 for 104 mouse TFs (Badis et al. 2009)). To assess the robustness of our results to this choice of affinity threshold, we also carried out a sensitivity analysis by varying τ within the interval (0.35, 0.45).

We consider 525 TFs from three kingdoms of life: animal, plant, and fungi (Table 1, Table S1). Specifically, we downloaded E-scores from the CIS-BP database for 86 TFs from *Mus musculus*, 217 TFs from *Arabidopsis thaliana*, and 118 TFs from *Neurospora crassa* (Weirauch et al. 2014). We downloaded E-scores of 104 additional *M. musculus* TFs from from the UniPROBE database (Badis et al. 2009; Newburger and Bulyk 2009). We chose to study these three species because they have more TFs characterized in the CIS-BP database than any other in their respective kingdoms. The TFs we study collectively represent 45 unique DNA binding domains, which can be thought of as distinct biophysical mechanisms by which TFs interact with DNA. A Venn diagram of the DNA binding domains in the three species is shown in Fig. S1A. In our dataset, several domains are common to all three species, whereas others are unique to one species. For example, *Homeodomain* TFs are found in all three species, but the family of *Zinc cluster* TFs is exclusive to *N. crassa*. This feature of our dataset provides an opportunity to discern whether the architecture of a GP map is governed by the peculiarities of particular binding domains or by the commonalities of TF-DNA interactions across binding domains. This is particularly useful in the context of TF-DNA interactions, because the set of TFs studied in a given species partly depends upon the interests of the field (e.g., cancer-associated TFs in humans vs. stress-responsive TFs in plants). By studying multiple species, we can ameliorate the potential effects of this bias.

A Venn diagram of the sites that bind TFs from the three species is shown in Fig. S1B. While many sites bind at least one TF in all three species (21.6%), many others bind TFs from just a single species. Specifically, 8.5%, 13.2% and 9.6% of sites uniquely bind TFs from *A. thaliana*, *M. musculus*, and *N. crassa*, respectively. The TFs bound by such sites do not preferentially belong to binding domains that

Table 1. Data analyzed in this study.

Species	Number of TFs	Number of DNA binding domains
Arabidopsis thaliana	217	25
Neurospora crassa	118	16
Mus musculus	190	25

are exclusive to a single species (Fig. S1C). In total, 14.4% of the 32,896 sites do not bind any of the TFs in our dataset.

In vivo data

We studied nucleotide diversity in putative TF binding sites in *Arabidopsis thaliana*. To do so, we gathered digital footprints from a DNase I hypersensitivity assay applied to root tissue (Sullivan et al. 2014). These data demarcate protein-bound open chromatin regions of the genome at single-nucleotide resolution, and can therefore be used to predict TF binding sites. We filtered the footprints to only include those that are at least eight nucleotides in length and that overlap the promoter regions of the 27,416 annotated protein-coding genes in the TAIR10 build of the *A. thaliana* genome, where a promoter is defined as the 500 bp upstream of a gene's transcription start site (Sullivan et al. 2014). This resulted in 123,330 footprints, which range in length from 8 to 40 bps (mean 16.56 bps). We used protein binding microarray data to determine whether the DNA sequence of each footprint has the potential to bind any of the 217 *A. thaliana* TFs. Specifically, for each TF, we determined whether the footprint contained a DNA sequence with an E-score > 0.35. If it did, we assigned the sequence to the TF as a putative binding site. If the footprint contained more than one binding site for a TF, we randomly chose one of the sites and assigned it to the TF. The number of binding sites thus assigned to a TF ranged from 142 for the TF LEC2 to 55,167 for the TF HMGA.

Nomenclature

We consider a *genotype space* of TF binding sites for each of the three species we study. This space comprises the set of all possible 32,896 double-stranded DNA sequences of length eight. The structure of this space can be described as a network, in which nodes represent TF binding sites and edges connect nodes if their corresponding sites differ by a single small mutation, specifically by a point mutation or by an indel (Payne and Wagner 2014). We refer to this network, which contains all possible genotypes, as

Ω. If two nodes are connected by an edge in Ω, we refer to them as *neighbors*.

Within this genotype space, we study a GP map in which each genotype is a DNA sequence (TF binding site), and the phenotype of this DNA sequence is its ability to bind one or more TFs (Payne and Wagner 2014). The set of genotypes with a particular phenotype is a *genotype set*. A single genotype may belong to multiple genotype sets, if the site binds multiple TFs. Each genotype set comprises one or more *genotype networks*, in which nodes are genotypes from the genotype set, and edges connect nodes that differ in a single small mutation, as in Ω. If a genotype set is fragmented into multiple genotype networks (connected components), it is usually the case that one network is much larger than the others (Payne and Wagner 2014; Aguilar-Rodríguez et al. 2017). We refer to this network as the *dominant genotype network* (Fig. S2).

Genotype networks are sub-networks of Ω, in which all genotypes have the same phenotype. We refer to mutations that do not change the phenotype as *neutral*, and to mutations that do change the phenotype as *non-neutral*. Thus, neutral mutations define the edges within genotype networks, whereas non-neutral mutations define the edges between genotype networks, or between a genotype network and unbound sequences. If two nodes are connected by an edge in a genotype network, we refer to them as *neutral neighbors*. We emphasize that we use the term "neutral" with respect to a specific phenotype, knowing full well that such mutations may not be neutral with respect to fitness.

Non-neutral mutations bridge the genotype networks of distinct phenotypes, thus helping to form the edges of a *phenotype network* (Alberch 1991). In such a network, each node represents the dominant genotype network of a specific TF, and edges connect nodes if (i) the associated genotype networks can be reached from one another by at least one non-neutral mutation, or (ii) these genotype networks share at least one genotype. In the latter case, we also say that the genotype networks *overlap*.

Genotype networks

To construct each genotype network of TF binding sites, we followed the same procedure as Payne and Wagner (2014). First, for each TF, we determined the set of sites that bind the TF (E-score > 0.35). Second, we used an alignment algorithm to calculate the mutational distance between all pairs of bound sites. Third, we used these mutational distances to define the edges of the genotype network by connecting two sites if they have a mutational distance of one.

We considered two kinds of mutations: point mutations, and indels that shift an entire, contiguous

binding site by one base (Fig. S3). Two DNA sequences of length eight can differ by a single point mutation in $3 \times 8 = 24$ different ways, because each of the sequence's nucleotides can mutate into any one of the three other nucleotides (Fig. S3A,B). In addition, there are $4 \times 2 = 8$ possible indels that can separate two DNA sequences of length eight. The reason is that the indels we consider can cause a shift in either the $5'$ or $3'$ direction, and in both cases the unaligned nucleotide can comprise any one of the four possible bases (Fig. S3C,D). There is therefore a maximum of $24 + 8 = 32$ single mutations that can separate two DNA sequences of length eight.

We determined the mutational distance between two DNA sequences using the Smith-Waterman alignment algorithm, prohibiting gaps in all alignments. For two sequences s_1 and s_2, we calculated the number of mismatches $m(s_1, s_2)$ and $m(s_1, s_2^{'})$, where $s_2^{'}$ is the reverse complement of s_2. We then took the minimum of $m(s_1, s_2)$ and $m(s_1, s_2^{'})$ as the mutational distance between s_1 and s_2.

Intra-network measures

We used several measures to characterize the internal structure of genotype networks (Newman 2010). The *diameter* of a genotype network is the longest of the shortest mutational paths between any pair of genotypes. The *characteristic path length* is the average of the shortest paths.

The *clustering coefficient* c measures the fraction of a genotype's neighbors that are also neighbors themselves, averaged across all genotypes in a genotype network (Watts and Strogatz 1998). Formally, the clustering coefficient is calculated as

$$c = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{2}{k_i(k_i - 1)} \sum_{j,k} A_{ij} A_{ik} A_{jk} \right), \tag{1}$$

where n is the number of genotypes, k_i is the degree of node i, A is the adjacency matrix of the genotype network, and j and k are the neighbors of node i.

The *degree assortativity* r of a genotype network measures the propensity for genotypes with a similar number of neighbors (i.e., vertex degree) to share an edge in a genotype network (Newman 2002). It corresponds to the Pearson correlation coefficient of the degrees of connected nodes, and therefore ranges in value from -1 to 1. When $r < 0$, the network is disassortative; when $r = 0$ it is uncorrelated; and when

$r > 0$, it is assortative. Assortativity is calculated as

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \tag{2}$$

where j_i and k_i are the degrees of the genotypes at the ends of the ith edge, and M is the number of edges in the genotype network.

The *route factor* q of a genotype network measures the average "directness" of the shortest mutational paths to a target genotype from all other genotypes in the network, relative to the shortest mutational paths to the target in Ω (the network used to describe genotype space). It is calculated as

$$q = \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{l_{i,\text{target}}}{d_{i,\text{target}}}, \tag{3}$$

where n is the number of nodes in the network, $l_{i,\text{target}}$ is the shortest mutational path between genotype i and the target genotype in the genotype network, and $d_{i,\text{target}}$ is the shortest mutational path between genotype i and the target genotype in Ω (Gastner and Newman 2006). We used the highest-affinity binding site as the target genotype. When $q = 1$, the genotype network is optimally distributed in Ω, in the sense that all paths to the target genotype are the shortest possible paths. When $q > 1$, the genotype network possesses paths to the target genotype that are longer than those in Ω, indicating deviations from an optimal distribution.

Inter-network measures

We characterized the arrangement of genotype networks in genotype space by measuring overlap and mutation probabilities ϕ_{qp} among all pairs of phenotypes q and p. We applied these measures at two levels of phenotypic granularity. In the first, the phenotype of a binding site genotype is its ability to bind one or more TFs. In the second, the phenotype of a binding site genotype is its ability to bind at least one TF in a class of TFs that share the same DNA binding domain. Regardless of the definition of phenotype, we applied the measures to the corresponding genotype networks in the same way.

The *overlap* $\mathcal{O}_{qp}$ of dominant genotype networks G_q and G_p, corresponding to phenotypes p and q, is

defined as

$$\mathcal{O}_{qp} = \frac{|S(G_q) \cap S(G_p)|}{|S(G_p)|}, \tag{4}$$

where $S(G_q)$ is the set of genotypes in genotype network G_q, and $|S(G_q)|$ is the number of genotypes in this set. Note that overlap is an asymmetric measure due to the normalization factor corresponding to the number of binding sites in G_p.

The fraction ϕ_{qp} of mutations to binding sites in genotype network G_p that create binding sites in genotype network G_q is defined as

$$\phi_{qp} = \frac{1}{|S(G_p)|} \sum_{i \in S(G_p)} \phi_q^{\text{local}}(i), \tag{5}$$

where

$$\phi_q^{\text{local}}(i) = \frac{n_i^q}{k_i}, \tag{6}$$

n_i^q is the number of neighbors of genotype i that have phenotype q, and k_i is the number of neighbors of genotype i in Ω. Thus, $\phi_q^{\text{local}}(i)$ is the fraction of genotype i's neighbors that have phenotype q. We used Eqs. 5 and 6 to calculate the mutational connectivity Φ_q of the genotype network of phenotype q from the genotype networks of all other phenotypes in genotype space as

$$\Phi_q = \sum_p \phi_{qp}. \tag{7}$$

The measure ϕ_{qp} is similar to the phenotypic accessibility A_{qp} of phenotype q from phenotype p, which is measured as

$$A_{qp} = \frac{|S(G_q) \cap \partial S(G_p)|}{|\partial S(G_p)|}, \tag{8}$$

where $S(G_q)$ is the set of genotypes in the dominant network of phenotype q and $\partial S(G_p)$ is the set of 1-mutant neighbors of the set $S(G_p)$ (Stadler et al. 2001; Cowperthwaite et al. 2008). We computed this measure as a point of comparison with ϕ_{qp}.

We complemented these global inter-network comparisons by comparing the phenotypic compositions of the local mutational neighborhoods of genotype pairs (i, j), using the Bhattacharyya coefficient

(Greenbury et al. 2016):

$$BC(i,j) = \sum_q \sqrt{\phi_q^{\text{local}}(i) \times \phi_q^{\text{local}}(j)}. \tag{9}$$

This coefficient quantifies the overlap of two distributions and therefore ranges from a minimum of zero when the phenotypic compositions of the mutational neighborhoods of genotypes i and j are maximally dissimilar to a maximum of one when they are identical. To quantify whether the phenotypic compositions of mutational neighborhoods are more similar among pairs of genotypes (i, j) that are neutral neighbors than among pairs of genotypes (i, k) that are not neutral neighbors, but are from the same genotype network, we computed the *similarity ratio* of the Bhattacharyya coefficients $BC(i, j)/BC(i, k)$. A ratio greater than 1 indicates that the phenotypic compositions of mutational neighborhoods of pairs of genotypes are more similar if those genotypes are connected by a neutral mutation than if they are not, and vice versa. Neighbors that are shared amongst genotypes i and j, and amongst i and k, were excluded from this analysis to provide a more conservative measure.

Null model

We compared our intra- and inter-network measures to those from GP maps contructed using the null model of Payne and Wagner (2014). Specifically, we randomly reassigned binding sites to TFs, such that the number of binding sites per TF did not change. We then constructed genotype networks from the reassigned binding sites, and calculated the intra- and inter-network measures. We repeated this process 1000 times for the GP maps of the three species that we study.

Shannon's diversity index

We assessed the amount of nucleotide diversity in each of a TF's putative binding sites in extant *A. thaliana* populations, using single nucleotide polymorphism data from the 1001 Genomes project (1001 Genomes Consortium 2016). Specifically, we calculated Shannon's diversity index D of each binding site as

$$D = -\frac{1}{8} \sum_{i=1}^{8} \sum_{j \in A,C,G,T} p_{ij} \log_2(p_{ij}), \tag{10}$$

where p_{ij} is the frequency of allele j at position i in the binding site. This measure takes on its minimum value of 0 when there is no diversity in the binding site. It takes on its maximum value of 2 when each of the four nucleotides occurs with equal frequency in all eight positions of the binding site.

Determining the number of TFs per DNA binding domain

We compared the number of TFs per DNA binding domain in our dataset to the same number in the proteomes of *A. thaliana* (UP000006548), *N. crassa* (UP000001805), and *M. musculus* (UP000000589), which we obtained from UniProt (The UniProt Consortium 2015). To find the number of proteins in each proteome with a match to a DNA binding domain, we employ the program hmmsearch from the software package HMMER (v3.1b2) (*http://hmmer.org/*). We used a cutoff of 0.01 for both the sequence e-value and the domain conditional e-value. We downloaded the hidden Markov models of each DNA-binding domain from the Pfam database (v27.0) (Finn et al. 2014).

Statement on reagent and data availability

All data used in this study are freely available in the UniPROBE and CIS-BP databases. Table S1 provides the necessary information to retrieve these data.

Results

Genotype space

We begin with a description of the network of all genotypes Ω, as this is the substrate of the genotype networks that we study in the subsequent sections. This network comprises 32,896 nodes and 523,728 edges. Its degree distribution is shown in Fig. S1D. The vast majority (96%) of genotypes have 32 neighbors, indicating that the network is nearly regular. The remaining 4% of sites possess peculiar features that are detailed in the Supplementary Material. Its diameter — the longest of the shortest paths between any two nodes — is eight, which corresponds to the maximum alignment distance between two sites. On average, however, pairs of TF binding sites are separated by only 4.385 mutations. The clustering coefficient of the network is 0.122, indicating that very few of a site's neighbors are neighbors themselves. This occurs because a site's neighbors can only be neighbors themselves if they differ in the

same nucleotide position. For example, the sequence ATATATAT has the neighbors ATATATAA and ATATATAG, which are neighbors themselves, but it also has neighbors such as CTATATAT and ACATATAT, which cannot be neighbors because they differ in two nucleotide positions. The network also lacks any meaningful assortativity by degree (indicated by an assortativity value of $r = 0.006$),which can be attributed in part to the low variance of the degree distribution.

Intra-network analyses

We first make some general observations about sets of genotypes that bind different TFs. The sizes of these genotype sets vary both within and across species, from a minimum of 2 sites for the *A. thaliana* TF Abf3 to 1,186 sites for the *M. musculus* TF Sp110. Across the three species, the average genotype set size is 374 sites. A total of 53% of these genotype sets comprise a single genotype network, whereas the remaining 47% comprise between 2 and 15 genotype networks. Despite such fragmentation, for 90% of the TFs, more than 95% of the genotype set belongs to the dominant genotype network (Table S1). We therefore carry out all of our analyses on the dominant genotype networks, as in previous work (Payne and Wagner 2014; Aguilar-Rodríguez et al. 2017). To simplify the presentation of our results, we focus on data from *M. musculus* in the main text, as it is representative of the data from *A. thaliana* and *N. crassa*, which we present in the Supplementary Material.

For the 190 *M. musculus* TFs, the average genotype network diameter is 6.7, varying from a minimum of 2 to a maximum of 14 (Fig. 1A; Table S1). In contrast, the characteristic path length — i.e., the average shortest distance between any pair of genotypes in a genotype network — is 3.2, less than half of the average network diameter (Fig. 1B; Table S1). These genotype networks are highly clustered, with an average clustering coefficient of 0.312 (Fig. 1C; Table S1). Taken together, the short characteristic path length relative to the diameter, and the high clustering coefficients, indicate that genotype networks of TF binding sites tend to fall within the family of "small world" networks (Watts and Strogatz 1998). These are networks that can be traversed in very few steps, like a random network, yet are highly clustered, like a regular lattice. In the context of TF binding sites, the "small world" property has two implications. First, it implies that binding sites are highly evolvable, because only very few mutations are required to travel across the network and potentially access new binding phenotypes. Second, it implies that binding sites are mutationally robust, because they may accumulate multiple mutations and still bind their cognate TF. Qualitatively similar results are obtained for the *A. thaliana* and *N. crassa* TFs (Figs.

S4A-C and S5A-C), indicating the consistency of these properties across three branches of the tree of life.

A recent numerical study suggests that assortativity (r) may influence evolutionary dynamics on genotype networks (Manrubia and Cuesta 2015). This measure, which ranges from $-1 \leq r \leq 1$, captures the propensity with which nodes of similar degree connect with one another (Newman 2002). Evolutionary dynamics on genotype networks that are assortative by degree ($r > 0$) may result in *phenotypic entrapment*, where the probability that an evolving population leaves a genotype network decreases with the time spent on it (Manrubia and Cuesta 2015). We find that most genotype networks exhibit a moderate amount of degree assortativity, possessing on average a value of $r = 0.25$ (Fig. 1D; Table S1). Degree assortativity is positively correlated with the size of the dominant genotype network (Spearman's $r = 0.57$, $p = 1.33 \times 10^{-17}$), such that disassortative genotype networks ($r < 0$) are always small (Fig. 1E). This likely reflects finite-size effects. Figs. S4D,E and S5D,E show that the same trends also exists in *A. thaliana* and *N. crassa* TFs. Assortativity is also positively correlated with characteristic path length (Spearman's $r = 0.56$, $p = 3.47 \times 10^{-19}$), indicating that as genotype networks become less "small-world", the potential for phenotypic entrapment increases. Finally, we emphasize that these trends in assortativity do not simply arise from the assortativity of Ω, because Ω shows very little assortativity ($r = 0.006$).

We next describe the structure of genotype networks using the *route factor* q. Fig. 1F shows the distribution of q for the dominant genotype networks of the 190 *M. musculus* TFs, where the target genotype is chosen to have the highest E-score. The distribution is heavily skewed toward $q = 1$, with an average route factor of $q = 1.01$ (Table S1). This indicates that genotype networks of TF binding sites are almost optimally distributed in Ω, meaning that almost all of the mutational paths in a genotype network that lead to the highest affinity sequence are the shortest possible mutational paths. Indeed, 38% of the genotype networks are optimally distributed, with $q = 1$. These results are consistent across the three species we study, as shown in Figs. S4F and S5F.

Figs. S6-S8 show that these intra-network statistics consistently differ from the null expectation. Specifically, the genotype networks constructued from the empirical data have longer diameters and characteristic path lengths, but shorter route factors, as well as higher clustering coefficients and assortativity values than the genotype networks contructed using the null model. Thus, the "small-worldness" of these genotype networks, as well as their efficient layout in genotype space, and assortative mixing patterns, are not expected structural properties according to the null model. Figs. S9-S11 show how

these intra-network statistics change as the binding affinity threshold is increased. In sum, they become more small-world (i.e., their characteristic path length decreases and their clustering coefficient increases) and slightly less assortative, with a more efficient layout in genotype space.

Finally, we ask what these intra-network measures tell us about the evolution of TF binding sites. In particular, we test a series of hypotheses about how the structural properties of genotype networks impact binding site diversity. To do so, we focus on *A. thaliana*, because two important sources of data are available for this species: Digital footprints from DNase I hypersensitivity assays (Sullivan et al. 2014), which can be used to predict TF binding sites, and high-quality single nucleotide polymorphism data (1001 Genomes Consortium 2016), which can be used to measure binding site diversity in extant populations. Our first hypothesis is that binding site diversity will increase as the number of binding sites in a genotype network increases. Our reasoning is that there are simply more sequences capable of binding a TF in a large genotype network than in a small genotype network, so these binding sites should exhibit more diversity. We find that this is indeed the case. The diversity of polymorphic TF binding sites exhibits a strong positive correlation with the size of the TF's genotype network (Fig. 2A; Spearman's correlation $\rho = 0.42, p = 1.71 \times 10^{-10}$). The second hypothesis is that binding site diversity will increase as the characteristic path length of a genotype network increases. Our reasoning is that genotype networks with large characteristic path lengths are more "spread out" in genotype space and will therefore permit the accumulation of more diversity than genotype networks with short characteristic path lengths. To test this hypothesis, we need to control for genotype network size, because this is positively correlated with characteristic path length (Spearman's correlation $\rho = 0.76, p = 4.46 \times 10^{-42}$). We find that even after controlling for genotype network size, binding site diversity increases with characteristic path length (Fig. 2B; Spearman's partial correlation $\rho = 0.23, p = 6.38 \times 10^{-4}$). The third hypothesis is that binding site diversity will increase as the route factor increases. Our reasoning is that genotype networks with high route factors are less "efficient" than those with short route factors, meaning that there are more sequences in the shortest paths between the highest-affinity sequence and other sequences in the network. This reduced efficiency should result in greater binding site diversity, which is indeed what we observe, even after controlling for genotype network size (Fig. 2C; Spearman's partial correlation $\rho = 0.21, p = 1.70 \times 10^{-3}$). The final hypothesis is that binding site diversity will decrease as assortativity increases. Our reasoning is that highly-connected nodes are "visited" more frequently by a population evolving on a genotype network (van Nimwegen et al. 1999), and in highly assortative networks, such nodes

preferentially connect to one another, which may make it difficult for an evolving population to escape the network's dense core, and thus lead to reduced diversity. However, after controlling for genotype network size, we find no evidence for this phenomenon (Spearman's partial correlation $p = 0.09$), possibly because the genotype networks we study exhibit limited variation in assortativity ($-0.47 \leq r \leq 0.51$ as compared to the full range of $-1 \leq r \leq 1$). Taken together, these analyses indicate that several of the structural properties of genotype networks affect the evolution of TF binding sites *in vivo*, particularly the extent to which binding sites accumulate genetic diversity.

Inter-network analyses

We now shift the scale of our analysis from local to global, transitioning from descriptions of individual genotype networks to descriptions of how these genotype networks overlap and interface with one another in Ω.

Overlap

Some TFs have similar binding preferences, especially if they are products of duplicate (paralogous) genes (Badis et al. 2009; Weirauch et al. 2014). The genotype networks of such TFs will therefore overlap, which has potential implications for TF crosstalk (Friedlander et al. 2016). Fig. 3A shows the overlap for all pairs (p, q) of TFs in the mouse dataset. Rows and columns correspond to individual TFs, and are arranged by DNA binding domain. The shading of matrix elements depicts overlap as the fraction of binding sites that are common to the genotype networks of two TFs. The matrix is asymmetric, because overlap is normalized by the genotype network size of TF q. Similar values of overlap are found in *A. thaliana* and *N. crassa* (Figs. S12A, S13A).

Paralogous TFs exhibit a high level of overlap in their genotype networks, as indicated by the block structure of the main diagonal in Fig. 3A. Even TFs with a *C2H2 ZF* binding domain, which exhibit the lowest levels of overlap, still share 9.14% of their binding sites on average. At the other end of the spectrum are two TFs with an *E2F* binding domain (E2F2 and E2F3), which share 92.73% of their binding sites. Overlap is not restricted to TFs from the same binding domain, as indicated by the blue shading off the main diagonal. For example, *ARID/BRIGHT* and *Sox* TFs share on average 16.5% of their binding sites. In fact, every single TF in the *M. musculus* dataset exhibits overlap in its genotype network with at least one other TF from a different binding domain. TFs with the same DNA binding

domain tend to share on average 27.2% of their binding sites, while TFs with different binding domains only share 1.88% on average (Wilcoxon rank-sum test, $p < 10^{-6}$). Fig. S14 compares genotype network overlap in the empirical data to the null model, and Fig. S15 shows that overlap gradually decreases as the binding affinity threshold increases. In sum, these results suggest that cognate and non-cognate TFs may often compete for the same binding sites, especially if the TFs are paralogs.

Interface

To characterize how the genotype networks of TF binding sites interface with one another, we calculated the fraction ϕ_{qp} of mutations to binding sites in the genotype network of TF p that create binding sites in the genotype network of TF q (Greenbury et al. 2016). The matrix in Fig. 3B shows ϕ_{qp} for all TFs in the mouse dataset. It is arranged as in Fig. 3A. Similar values of ϕ_{qp} are found in *A. thaliana* and *N. crassa* (Figs. S12B, S13B).

Of the 35,910 pairwise comparisons depicted in Fig. 3B, 31,548 (87.9%) have $\phi_{qp} > 0$ (as compared to 33.2% in the null model). This means that genotype networks of TF binding sites interface with one another to such an extent that it is usually possible to evolve at least one of a TF's binding sites via a single small mutation to a binding site of nearly any other TF. On average, the ϕ_{qp} between the genotype networks of TFs with the same binding domain is higher than that of TFs with different binding domains (0.139 compared to 0.016; Wilcoxon rank-sum test, $p < 10^{-6}$). Fig. S17 compares ϕ_{qp} in the empirical data to that of the null model, and Fig. S18 shows how ϕ_{qp} decreases as the binding affinity threshold increases.

Some pairs of TFs with different binding domains have high ϕ_{qp}. For example, the genotype networks of TFs with a *SAND* binding domain have a higher ϕ_{qp}, on average, with the genotype networks of TFs with a *bZIP* binding domain than they do with the genotype networks of TFs with the same binding domain. To investigate this further, we compare ϕ_{qp} to the null expectation (Payne and Wagner 2014), which is equivalent to the fraction f_q of genotypes with phenotype q (Schaper and Louis 2014; Greenbury et al. 2016). We consider the TF Hes7 as phenotype q, because it has the largest genotype network of the TFs with a bHLH binding domain, which is the domain with the largest number of TFs in our dataset. We find that the null model does not provide a reasonable approximation to the empirical data (Fig. S16), in contrast with earlier observations in computational models of GP maps (Greenbury et al. 2016). This means that the overall frequency of a phenotype — i.e., the fraction of genotypes with that phenotype

— is not a good indicator of the probability that a randomly chosen non-neutral mutation leads to that phenotype. We find that for TFs with the same binding domain as the focal TF Hes7, ϕ_{qp} is typically larger than the null expectation (Fig. S16, filled circles). Since such TFs often bind similar sets of sites (Weirauch et al. 2014), this observation corroborates the intuition that their genotype networks interface more than expected by chance. However, such TFs do not fully account for the observed deviation from the null model, because removing them from the linear fit of ϕ_{qp} to f_q barely improves the coefficient of determination ($R^2 = 0.1337$, as compared to $R^2 = 0.082$). In sum, the arrangement of genotype networks in Ω deviates substantially from the null expectation, and this deviation is not explained by TF paralogs binding similar sets of sequences; even the arrangement of genotype networks of non-paralogous TFs deviates from the null expectation.

GP maps often exhibit correlations in their local mutational neighborhoods (Greenbury et al. 2016). We therefore sought to determine if the composition of such neighborhoods — in terms of the phenotypes that occur in them — might deviate from the null expectation. To do so, we compared the composition of the mutational neighborhoods of pairs of neighboring genotypes on a genotype network to the mutational neighborhoods of randomly selected pairs of non-neighboring genotypes from the same genotype network, removing neighbors that are shared by the genotypes being compared. We used this comparison to compute a similarity ratio that is greater than unity when neighboring genotypes have more similar sets of phenotypes in their mutational neighborhoods than do non-neighboring genotypes (Greenbury et al. 2016). Fig. S19 shows a histogram of this similarity ratio for all possible pairs of neighboring genotypes in the genotype network for the mouse TF Sp110, which we have chosen to exemplify this result because it has the largest genotype network in the *M. musculus* dataset. The mean is 1.465 ± 0.006, which deviates significantly from the null expectation of unity (one-sample t-test, $t = 79.87$, $p < 10^{-6}$). We made similar observations in the *A. thaliana* and *N. crassa* data (Figs. S20, S21). Fig. S22 shows that the similarity ratio is higher than expected under the null model, and Fig. S23 shows that although the similarity ratio decreases as the binding affinity threshold increases, it always remains above unity.

So far, we have only considered how genotype networks interface with one another. Since mutations that abrogate TF binding are also important for regulatory evolution (Guerreiro et al. 2013), we now turn our attention to the interface of genotype networks with the regions of Ω that do not bind any TF. Such unbound regions are not small: They comprise 51%, 48%, and 39% of Ω in *A. thaliana*, *N. crassa*, and *M. musculus*, respectively. For each TF p, we calculate the fraction $\phi_{\mathrm{unbound},p}$ of mutations to binding

sites in the genotype network for TF p that create unbound sites — i.e., sites that do not bind any TF in our dataset, for the respective species. We then divide this number by the fraction f_{unbound} of unbound sites, which is the null expectation for $\phi_{\text{unbound},p}$ (Greenbury et al. 2016). Thus, this ratio will equal unity when the empirical data is well represented by the null model. Fig. S24 shows this ratio for all of the mouse TFs. It is consistently below unity. This indicates that unbound sites occur less frequently in the mutational neighborhoods of bound sites than is expected under the null model. Thus, the interface of genotype networks with unbound sites in Ω is qualitatively different from the interface of genotype networks with one another. We made similar observations in the *A. thaliana* and *N. crassa* data (Figs. S25, S26), and these findings are insensitive to the binding affinity threshold (Fig. S27).

Finally, we sum across the columns of Fig. 3B to obtain a global measure Φ_q of the mutational connectivity of the genotype network of phenotype q with the genotype networks of all other phenotypes in genotype space. This measure is related to, and highly correlated with, a popular measure called *phenotypic accessibility* (Stadler et al. 2001; Cowperthwaite et al. 2008) (Spearman's $r = 0.95$, $p < 10^{-6}$; Fig. S28). The main difference is that Φ_q accounts for genotype network overlap. We find that Φ_q increases with genotype network size (Spearman's $r = 0.64$, $p < 10^{-6}$; Fig. S29), indicating that non-neutral mutations to TF binding sites are more likely to create binding sites for low-specificity TFs than for high-specificity TFs, because low-specificity TFs have larger genotype networks (Payne and Wagner 2014). We also find that Φ_q increases with genotype network size in *A. thaliana* (Fig. S30) and *N. crassa* (Fig. S31).

Phenotype space covering

To further characterize how genotype networks of TF binding sites overlap and interface with one another, we calculated the average fraction of phenotypes found within n mutations of each binding site, for each TF. We refer to this measure, which has been introduced in a different context as *shape space covering* (Schuster et al. 1994), as *phenotype space covering*, and we call a phenotype "covered" if it is found within a mutational radius of a genotype. We again use the mouse TF Sp110 to exemplify our findings.

We consider two variants of phenotype space covering. In the first, we determine the phenotypes of all genotypes within a mutational radius of n, such that all mutations are neutral (i.e., the binding sites are part of the same genotype network). This analysis is therefore a further characterization of genotype network overlap. We find for the murine TF Sp110 that within just a single mutation ($n = 1$), an average

of 8.51% of the phenotypes are covered, and that within a mutational radius of $n > 4$, a total of 46.31% of the phenotypes are covered (Fig. 4A). The genotype network for Sp110 therefore overlaps with the genotype networks of nearly half of the mouse TFs in our dataset. Fig. S32A shows that this does not happen under the null model where phenotype space covering is close to zero for all n. We then asked how the maximum proportion of phenotypes covered (e.g., 46.31% for Sp110) relates to the size of a genotype network. Fig. 4B shows that this maximum proportion is largely determined by the size of the dominant genotype network (Spearman's $r = 0.76$, $p < 10^{-6}$), such that larger dominant genotype networks cover more phenotypes. Figs. S33A - S35A show how this maximum proportion decreases as the binding affinity threshold increases.

In the second variant of phenotype space covering, we consider all genotypes within a mutational radius of n, such that all mutations are non-neutral. The proportion of phenotypes covered within a mutational radius of $n = 1$ does not differ from the first variant, but it increases more rapidly with n, such that all phenotypes are covered within a mutational radius of $n > 4$ (Fig. 4A). Fig. S32B shows that this increase is reduced under the null model. Moreover, there is no variation in this measure when $n > 4$, meaning that all phenotypes are covered within this mutational radius from any binding site of Sp110. Across all of the mouse TFs, $n = 4.5$ is the average mutational radius for which the coefficient of variance (σ/μ) in the proportion of phenotypes covered becomes smaller than 1%. There are 33 TFs that cover more than 99% of all phenotypes within a radius of $n \leq 4$. Remarkably, 5 of these networks are extremely small, comprising between 8 and 11 binding sites (TFs Arnt2, Fosl1, Hes2, Jun, and Olig3). These binding repertoires are therefore exceptionally evolvable. Figs. S33B - S35B show how the radius at which all phenotypes are covered increases as the binding affinity threshold increases.

Genotype networks of DNA binding domains

The GP map we study can be analyzed at multiple levels of granularity. We have so far considered a fine-grained analysis, in which each genotype is a DNA sequence, and the phenotype is the sequence's ability to bind one or more TFs. We now consider a more coarse-grained analysis, in which each genotype is a DNA sequence, and the phenotype is the sequence's ability to bind at least one TF in a class of TFs that share the same DNA binding domain. Studying the overlap and interface of such genotype networks complements our previous analyses by describing how TFs with different binding domains may compete for the same sites, and how DNA mutations may transfer regulatory control from a TF with one DNA

binding domain to a TF with a different binding domain.

Fig. 5A shows the extent of overlap among all pairs of genotype networks for the 25 DNA-binding domains in the *M. musculus* dataset. Such overlap is pervasive. For example, there are six binding domains with genotype networks that overlap the genotype networks of every other binding domain in the dataset (*bHLH*, *bZIP*, *C2H2 ZFs*, *Ets*, *Homeodomain*, *SAND*). Even the *AP-2* and *Ndt80/PhoG* binding domains, which exhibit the lowest levels of overlap, still overlap with 14 (56%) of the other domains. In total, 504 of the 600 pairs of binding domains exhibit overlap in their genotype networks. It is therefore common for TFs with different binding domains to recognize some of the same sites, further highlighting the potential for crosstalk in transcriptional regulation (Friedlander et al. 2016). Similar patterns hold in *A. thaliana* and *N. crassa* (Figs. S36A, S37A), even though these species have several binding domains that are not present in the *M. musculus* dataset. Such overlap therefore appears to be a general consequence of the low specificity with which eukaryotic TFs interface with DNA, rather than a consequence of the binding preferences of any particular binding domain.

Fig. 5B shows ϕ_{qp} for all pairs of the 25 DNA-binding domains in the *M. musculus* dataset. As with overlap, we observe an increase in ϕ_{qp} as we shift the level of analysis from TFs to DNA binding domains. A total of 590 (98.3%) binding domain pairs exhibit non-zero ϕ_{qp}. Mutations in TF binding sites could thus commonly transfer regulatory control among TFs with different binding domains. Similar observations are made for *A. thaliana* and *N. crassa* (Figs. S36B, S37B). We also studied how the different genotype networks of DNA binding domains interface with one another through the visualization of phenotype networks (Figs. S38-S40).

Finally, Φ_q scales with genotype network size at the level of DNA binding domains, just as it did at the level of individual TFs (compare Figs. S41 - S43 with Figs. S29-S31). However, since the number of TFs per binding domain in the *M. musculus* dataset also scales with genotype network size (Fig. S44A), we were concerned that these trends may stem from ascertainment bias. This could occur if the number of TFs per binding domain in the *M. musculus* dataset was not representative of the number of TFs per binding domain in the *M. musculus* genome. Figs. S44B,C show that this is not the case. Both the number of TFs per binding domain in the *M. musculus* dataset, and the size of the corresponding genotype network, scale with the number of TFs per binding domain in the *M. musculus* genome. We made similar observations in *A. thaliana* and *N. crassa* (Figs. S45, S46).

Discussion

The concept of a genotype-phenotype (GP) map can be traced back to the work of Sewall Wright (Wright 1932), Conrad H. Waddington (Waddington 1959), and John Maynard Smith (Maynard Smith 1970). However, the term GP map ("genotype-phenotype mapping") was only coined in 1970 by Jim Burns (Burns 1970), who recognized the importance of incorporating a mechanistic perspective into the evolutionary framework of population genetics, thus outlining the research programme that has come to be known as evolutionary systems biology (Soyer and O'Malley 2013). The term was re-introduced in 1991 by the developmental biologist Pere Alberch (Alberch 1991), who was interested in macroscopic phenotypes arising from complex developmental processes. The study of GP maps is currently shifting away from the conceptual and computational models that shaped the thinking of the founders of the field, toward empirical data derived from high-throughput assays (Rowe et al. 2010; Fowler et al. 2010; Hinkley et al. 2011; Jiménez et al. 2013; Melamed et al. 2013; Buenrostro et al. 2014; Findlay et al. 2014; Olson et al. 2014; Podgornaia and Laub 2015; Julien et al. 2016; Li et al. 2016; Puchta et al. 2016; Qiu et al. 2016; Sarkisyan et al. 2016). Our study is part of this shift. We have used experimental data from protein binding microarrays to analyze the architecture of an empirical GP map, in which each genotype is a short DNA sequence, and the phenotype of the sequence is its ability to bind one or more TFs. This study expands upon previous analyses of this map (Payne and Wagner 2014; Aguilar-Rodríguez et al. 2017) by providing more nuanced descriptions of individual genotype networks, detailed characterizations of how these networks overlap and interface with one another, and does so at two levels of phenotypic granularity.

Our analyses of individual genotype networks provide two new insights into their structure. First, they tend to be "small-world"(Watts and Strogatz 1998), an observation that furthers our understanding of the "robust-yet-evolvable" nature of TF binding sites (Payne and Wagner 2014): While binding sites tend to be highly clustered in their genotype network (robustness), it remains possible to traverse the network with just a few mutations, thus providing efficient access to adjacent genotype networks (evolvability). Indeed, the route factor of these genotype networks indicates that they are almost optimally distributed in genotype space, in the sense that almost all genotypes are connected to a central target genotype through the shortest mutational paths. These structural properties have implications for the evolution of TF binding sites. Specifically, we find that they affect the accumulation of genetic diversity in extant

populations of *Arabidopsis thaliana*, such that binding site diversity increases as a genotype network's characteristic path length or route factor increases.

The second new insight is that genotype networks are assortative, meaning that robust (i.e., highly connected) binding sites are likely to neighbor other robust binding sites. The potential implication for the evolution of TF binding sites is reduced diversity, because an evolving population tends to accumulate in such densely connected regions of genotype networks (van Nimwegen et al. 1999), which may lead to "phenotypic entrapment" (Manrubia and Cuesta 2015), a phenomenon in which an evolving population becomes less likely to leave its genotype network the more time it spends on it. However, we do not find evidence of this in our analysis of polymorphism data from *A. thaliana*, inasmuch as we do not find a significant relationship between assortativity and binding site diversity. One possible explanation is that the genotype networks we study exhibit limited variation in assortativity.

A sometimes underappreciated feature of GP maps is that genotypes may have more than one phenotype (West-Eberhard 2003; Wagner and Zhang 2011), which means that genotype networks may overlap. Even if we restrict our examples to the molecular realm, they are numerous: An RNA transcript can be translated into different proteins (Bratulic et al. 2015), an amino acid sequence can fold into different conformational structures (Bloom et al. 2006), and a promiscuous enzyme can catalyze different reactions (Nam et al. 2012). In the GP map studied here, such overlap is also pervasive, both at the level of individual TFs and of DNA binding domains. It implies competition for binding sites among cognate and non-cognate TFs, a phenomenon known as "crosstalk." Recent modeling work suggests that crosstalk is an inevitable feature of transcriptional regulation in species that employ limited-specificity TFs (Friedlander et al. 2016), such as the three eukaryotic species studied here. This is important because crosstalk places constraints on the function and evolution of transcriptional regulatory networks. Our results provide an empirical complement to these earlier theoretical findings, by providing estimates of how much crosstalk can occur among TFs and binding domains. However, it is worth highlighting that these estimates are based on *in vitro* measurements of TF binding preferences. The myriad complexities of *in vivo* TF-DNA interactions (Siggers and Gordân 2014), including epigenetic marks, local sequence and chromatin context, as well as interactions with protein partners, will certainly affect these estimates. Our ability to interrogate the effects of these complexities on TF-DNA interactions is continuing to advance (Hu et al. 2013; Levo et al. 2015; Isakova et al. 2017; Levo et al. 2017), and we believe that genotype networks will provide a useful framework for studying how such complexities mitigate crosstalk in tran-

scriptional regulation. Of particular interest is the role played by chromatin silencing (Beisel and Paro 2011), which may mitigate crosstalk by making binding sites unavailable in the presence of non-cognate TFs.

Our analysis of how genotype networks interface with one another has implications for the emergence of evolutionary innovations, because mutations in cis-regulatory regions may produce novel gene expression patterns (Wray 2007; Prudhomme et al. 2007). In particular, single-base pair mutations in TF binding sites can shift the regulatory control of a gene from one TF to another, and this may cause profound phenotypic change. For example, such mutations led to the differential expression of Rhodopsin genes in different subsets of *Drosophila* photoreceptors (Rister et al. 2015), which facilitated the discrimination of a wide spectrum of optical stimuli, and thus drastically changed how flies perceive their environment. In the GP map studied here, it has been previously shown that genotype networks are so intertwined that it is usually possible to mutate at least one of a TF's binding sites to a binding site of nearly any other TF (Payne and Wagner 2014). This means that mutation can readily shift the regulatory control of a gene from one TF to another, a shift that may lead to an adaptive change in gene expression. Here, we provide a more detailed and nuanced view of TF binding site evolvability. At the most local scale, evolvability is relatively low because neutral neighbors tend to have highly similar mutational neighborhoods, which decreases the diversity of novel phenotypes that may arise via a single point mutation to any one binding site (Greenbury et al. 2016). However, only very few mutations are required to shift regulatory control from the cognate TF to nearly any other TF in our dataset. At even this intermediate scale, TF binding sites are therefore remarkably evolvable.

An important challenge in the biological sciences is to provide a comprehensive description of the architecture of an empirical GP map. The hyper-astronomical size of genotype space renders this challenge impossible for most biological systems of interest, including macromolecules, regulatory circuits, and metabolisms (Louis 2016). Even for the relatively small genotype space studied here, we fall short of a comprehensive description. The reason is that we do not have data describing the binding preferences of every TF from each of our three study species. However, the data we do have are a representative sampling of each species' TF repertoire. There are two reasons for this. The first is that the assayed TFs were intentionally selected to exhibit an even balance among DNA binding domains and to survey different levels of sequence similarity (Weirauch et al. 2014). The second is that the number of TFs per binding domain in our dataset is correlated with the number of TFs per binding domain in the genomes of

each species (The UniProt Consortium 2015). This study therefore provides a high-resolution depiction of the architecture of an empirical GP map.

References

1001 Genomes Consortium. 2016. 1,135 genomes reveal the global pattern of polymorphism in *arabidopsis thaliana*. Cell 166:1–11.

Aguilar-Rodríguez, J., J. L. Payne, and A. Wagner. 2017. A thousand adaptive landscapes and their navigability. Nat. Ecol. Evol. 1:0045.

Aguirre, J., J. Buldú, M. Stich, and S. Manrubia. 2011. Topological structure of the space of phenotypes: the case of RNA neutral networks. PLOS ONE 6:e26324.

Ahnert, S. E. 2017. Structural properties of genotype-phenotype maps. J. R. Soc. Interface 14:20170275.

Alberch, P. 1991. From genes to phenotype: dynamical systems and evolvability. Genetica 84:5–11.

Badis, G., M. F. Berger, A. A. Philippakis, S. Talukder, A. R. Gehrke, S. A. Jaeger, E. T. Chan, G. Metzler, A. Vedenko, X. Chen, H. Kuznetsov, C. Wang, D. Coburn, D. E. Newburger, Q. Morris, T. R. Hughes, and M. L. Bulyk. 2009. Diversity and complexity in DNA recognition by transcription factors. Science 324:1720–1723.

Barve, A. and A. Wagner. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature 500:203–206.

Beisel, C. and R. Paro. 2011. Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. 12:123–135.

Berger, M. F., A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep, and M. L. Bulyk. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. 24:1429–1435.

Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA 103:5896–5874.

Bornberg-Bauer, E. and H. Chan. 1999. Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. Proc. Nat. Acad. Sci. USA 96:10689–10694.

Bratulic, S., F. Gerber, and A. Wagner. 2015. Mistranslation drives the evolution of robustness in TEM-1 β-lactamase. Proc. Natl. Acad. Sci. USA 112:12758–12763.

Buenrostro, J. D., C. L. Araya, L. M. Chircus, C. J. Layton, H. Y. Chang, M. P. Snyder, and W. J. Greenleaf. 2014. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat. Genet. 32:562–568.

Burns, J., 1970. The synthetic problem and the genotype-phenotype relation in cellular metabolism. Pp. 47–51, *in* C. H. Waddington, ed. Towards a Theoretical Biology. Volume 3: Drafts. An IUBS Symposium. Aldine Publishing Company, Chicago.

Cartwright, R. 2009. Problems and solutions for estimating indel rates and length distributions. Mol. Biol. Evol. 26:473–480.

Chen, J.-Q., Y. Wu, H. Yang, J. Bergelson, M. Kreitman, and D. Tian. 2009. Variation in the ratio of nucleotide subsitution and indel rates across genomes in mammals and bacteria. Mol. Biol. Evol. 26:1523–1531.

Ciliberti, S., O. C. Martin, and A. Wagner. 2007. Innovation and robustness in complex regulatory gene networks. Proc. Natl. Acad. Sci. USA 104:13591–13596.

Cotterell, J. and J. Sharpe. 2010. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol. Syst. Biol. 6:425.

Cowperthwaite, M. C., E. P. Economo, W. R. Harcombe, E. L. Miller, and L. A. Meyers. 2008. The ascent of the abundant: How mutational networks constrain evolution. PLOS Comput. Biol. 4:e1000110.

Draghi, J. A., T. L. Parsons, G. P. Wagner, and J. B. Plotkin. 2010. Mutational robustness can facilitate adaptation. Nature 463:353–355.

Ferrada, E. 2014. The amino acid alphabet and the architecture of the protein sequence-structure map. I. Binary alphabets. PLOS Computat. Biol. 10:e1003946.

Findlay, G. M., E. A. Boyle, R. J. Hause, J. C. Klein, and J. Shendure. 2014. Saturation editing of genomic regions by multiplex homology-directed repair. Nature 513:120–123.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, and R. Y. Eberhardt et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222–D230.

Fortuna, M., L. Zaman, C. Ofria, and A. Wagner. 2017. The genotype-phenotype map of an evolving digital organism. PLOS Computat. Biol. 13:e1005414.

Fowler, D. M., C. L. Araya, S. J. Fleishman, E. H. Kellog, and J. J. Stephany et al. 2010. High-resolution mapping of protein sequence-function relationships. Nat. Meth. 7:741–746.

Friedlander, T., R. Prizak, C. C. Guet, N. H. Barton, and G. Tkačik. 2016. Intrinsic limits to gene regulation by global crosstalk. Nat. Comm. 7:12307.

Gastner, M. T. and M. E. J. Newman. 2006. Shape and efficiency in spatial distribution networks. J. Stat. Mech. P. P01015.

Greenbury, S. F., I. G. Johnson, A. A. Louis, and S. E. Ahnert. 2014. A tractable genotype-phenotype map modelling the self-assembly of protein quaternary structure. J. R. Soc. Interface 11:20140249.

Greenbury, S. F., S. Schaper, S. E. Ahnert, and A. A. Louis. 2016. Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. PLOS Computat. Biol. 12:e1004773.

Guerreiro, I., A. Nunes, J. M. Woltering, A. Casaca, A. Nóvoa, T. Vinagre, M. E. Hunter, D. Duboule, and M. Mallo. 2013. Role of a polymorphism in a hox/pax-responsive enhancer in the evolution of the vertebrate spine. Proc. Natl. Acad. Sci. USA 110:10682–10686.

Hinkley, T., J. Martins, C. Chappey, M. Haddad, and E. Stawiski et al. 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat. Genet. 43:487–489.

Holland, P. W., K. B. Laskey, and S. Leinhardt. 1983. Stochastic blockmodels: First steps. Soc. Networks 5:109–137.

Hu, S., J. Wan, Y. Su, Q. Song, Y. Zeng, H. N. Nguyen, J. Shin, E. Cox, H. S. Rho, C. Woodard, S. Xia, S. Liu, H. Lyu, G.-L. Ming, H. Wade, H. Song, J. Qian, and H. Zhu. 2013. DNA methylation presents distinct binding sites for human transcription factors. eLife 2:e00726.

Hu, T., J. L. Payne, W. Banzhaf, and J. H. Moore. 2012. Evolutionary dynamics on multiple scales: a quantitative analysis of the interplay between genotype, phenotype, and fitness in linear genetic programming. Genet. Program Evol. M. 13:305–337.

Isakova, A., R. Groux, M. Imbeault, P. Rainer, D. Alpern, R. Dainese, G. Ambrosini, D. Trono, P. Bucher, and B. Deplancke. 2017. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. Nat. Meth. 14:316–322.

Jiménez, J. I., R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod, and I. A. Chen. 2013. Comprehensive experimental fitness landscape and evolutionary network for small RNA. Proc. Natl. Acad. Sci. USA 110:14984–14989.

Julien, P., B. Miñana, P. Baeza-Centurion, J. Valcárcel, and B. Lehner. 2016. The complete local genotype-phenotype landscape for the alternative splicing of a human exon. Nat. Comm. 7:11558.

Lehner, B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. 14:168–178.

Levo, M., T. Avnit-Sagi, M. Lotan-Pompan, Y. Kalma, A. Weinberger, Z. Yakhini, and E. Segal. 2017. Systematic investigation of transcription factor activity in the context of chromatin using massively parallel binding and expression assays. Mol. Cell 65:604–617.

Levo, M., E. Zalckvar, E. Sharon, A. C. D. Machado, Y. Kalma, M. Lotam-Pompan, A. Weinberger, Z. Yakhini, R. Rohs, and E. Segal. 2015. Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. 25:1018–1029.

Li, C., W. Qian, C. J. Maclean, and J. Zhang. 2016. The fitness landscape of a tRNA gene. Science 352:837–840.

Lipman, D. J. and J. W. Wilbur. 1991. Modelling neutral and selective evolution of protein folding. Proc. R. Soc. London Ser. B 245:7–11.

Louis, A. A. 2016. Contingency, convergence and hyper-astronomical numbers in biological evolution. Stud. Hist. Philos. Biol. Biomed. Sci. 58:107–116.

Manrubia, S. and J. A. Cuesta. 2015. Evolution on neutral networks accelerates the ticking rate of the molecular clock. J. R. Soc. Interface 12:20141010.

Maynard Smith, J. 1970. Natural selection and the concept of a protein space. Nature 225:563–564.

McCandlish, D. M. 2013. On the findability of genotypes. Evolution 67:2592–2603.

Melamed, D., D. L. Young, C. E. Gamble, C. R. Miller, and S. Fields. 2013. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. RNA 19:1537–1551.

Nakagawa, S., S. S. Gisselbrecht, J. M. Rogers, D. L. Hartl, and M. L. Bulyk. 2013. Dna-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. USA 110:12349–12354.

Nam, H., N. E. Lewis, J. A. Lerman, D.-H. Lee, R. L. Chang, D. Kim, and B. O. Palsson. 2012. Network context and selection in the evolution to enzyme specificity. Science 337:1101–1104.

Newburger, D. E. and M. L. Bulyk. 2009. Uniprobe: an online database of protein binding microarray data on protein-dna interactions. Nucleic Acids Res. 37:D77–82.

Newman, M., 2010. Networks: An Introduction. Oxford University Press, Oxford.

Newman, M. E. J. 2002. Assortative mixing in networks. Phys. Rev. Lett. 89:208701.

———. 2016. Equivalence between modularity optimization and maximum likelihood methods for community detection. Phys. Rev. E 94:052315.

Nowicki, K. and T. A. B. Snijders. 2001. Estimation and prediction for stochastic blockstructures. J. Amer. Statist. Assoc. 96:1077–1087.

Olson, C. A., N. C. Wu, and R. Sun. 2014. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr. Biol. 24:2643–2651.

Payne, J. L. and A. Wagner. 2013. Constraint and contingency in multifunctional gene regulatory circuits. PLOS Comput. Biol. 9:e1003071.

———. 2014. The robustness and evolvability of transcription factor binding sites. Science 466:714–719.

Peel, L., D. B. Larremore, and A. Clauset. 2017. The ground truth about metadata and community detection in networks. Sci. Adv. 3:e1602548.

Peixoto, T. P. 2014. Efficient monte carlo and greedy heuristic for the inference of stochastic block models. Phys. Rev. E 89:012804.

Pigliucci, M. 2010. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Proc. R. Soc. London Ser. B 365:557–566.

Podgornaia, A. I. and M. T. Laub. 2015. Pervasive degeneracy and epistasis in a protein-protein interface. Science 347:673–677.

Prudhomme, B., N. Gompel, and S. B. Carroll. 2007. Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. USA 104:8605–8612.

Ptashne, M. and A. Gann, 2002. Genes & Signals. Cold Spring Harbor Laboratory Press, New York.

Puchta, O., B. Cseke, H. Czaja, D. Tollervey, G. Sanguinetti, and G. Kudla. 2016. Network of epistatic interactions within a yeast snoRNA. Science 352:840–844.

Qiu, C., O. C. Erinne, J. M. Dave, P. Cui, H. Jin, N. Muthukrishnan, L. K. Tang, S. G. Babu, K. C. Lam, P. J. Vandeventer, R. Strohner, J. Van den Brulle, S.-H. Sze, and C. D. Kaplan. 2016. High-resolution phenotypic landscape of the RNA Polymerase II Trigger Loop. PLOS Genet. 12:e1006321.

Raman, K. and A. Wagner. 2011. The evolvability of programmable hardware. J. R. Soc. Interface 8:269–281.

Rister, J., A. Razzaq, P. Boodram, N. Desai, C. Tsanis, H. Chen, D. Jukam, and C. Desplan. 2015. Single-base pair differences in a shared motif determine differential *rhodopsin* expression. Science 350:1258–1261.

Rodrigues, J. F. M. and A. Wagner. 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. PLOS Computat. Biol. 5:e1000613.

Rohs, R., S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig. 2009. The role of DNA shape in protein-DNA recognition. Nature 461:1248–1253.

Rowe, W., M. Platt, D. C. Wedge, P. J. Day, D. B. Kell, and J. Knowles. 2010. Analysis of a complete DNA-protein affinity landscape. J. R. Soc. Interface 7:397–408.

Salazar-Ciudad, I. and M. Marín-Riera. 2013. Adaptive dynamics under development-based genotype-phenotype maps. Nature 497:361–364.

Sarkisyan, K. S., D. A. Bolotin, M. V. Veer, D. R. Usmanova, and A. S. Mishin et al. 2016. Local fitness landscape of the green fluorescent protein. Nature 533:397–401.

Schaper, S. and A. A. Louis. 2014. The arrival of the frequent: How bias in genotype-phenotype maps can steer populations to local optima. PLOS ONE 9:e86635.

Schuster, P., W. Fontana, P. F. Stadler, and I. L. Hofacker. 1994. From sequences to shapes and back: a case study in RNA secondary structures. Proc. R. Soc. London Ser. B 255:279–284.

Siggers, T. and R. Gordân. 2014. Protein-DNA binding: complexities and multi-protein codes. Nucleic Acids Res. 42:2099–2111.

Soyer, O. and M. O'Malley. 2013. Evolutionary systems biology: What it is and why it matters. BioEssays 35:696–705.

Stadler, B. M. R., P. F. Stadler, G. P. Wagner, and W. Fontana. 2001. The topology of the possible: Formal spaces underlying patterns of evolutionary change. J. Theor. Biol. 213:241–274.

Sullivan, A. M., A. A. Arsovski, J. Lempe, K. L. Bubb, and M. W. Weirauch et al. 2014. Mapping and dynamics of regulatory DNA and transcription factor networks in *a. thaliana*. Cell Rep. 8:2015–2030.

The UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.

van Nimwegen, E., J. P. Crutchfield, and M. Huynen. 1999. Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA 96:9716–9720.

Waddington, C. H., 1959. The Strategy of the Genes. Macmillan, New York, US.

Wagner, A. 2008. Robustness and evolvability: a paradox resolved. Proc. R. Soc. London Ser. B 275:91–100.

———. 2014. Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity. Biophys. J. 106:955–965.

Wagner, G. P. and J. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. 12:204–213.

Watts, D. J. and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393:440–442.

Weirauch, M. T., A. Yang, M. Albu, A. G. Cote, and A. Montenegro-Montero et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158:1431–1443.

West-Eberhard, M. J., 2003. Developmental Plasticity and Evolution. Oxford University Press, Oxford, UK.

Wray, G. A. 2007. The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. 8:206–216.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *in* D. F. Jones, ed. Proceedings of the Sixth International Congress on Genetics, vol. 1, Pp. 356–366. The Genetics Society of America.

Zhang, X., T. Martin, and M. Newman. 2015. Identification of core-periphery structure in networks. Phys. Rev. E 91:032803.

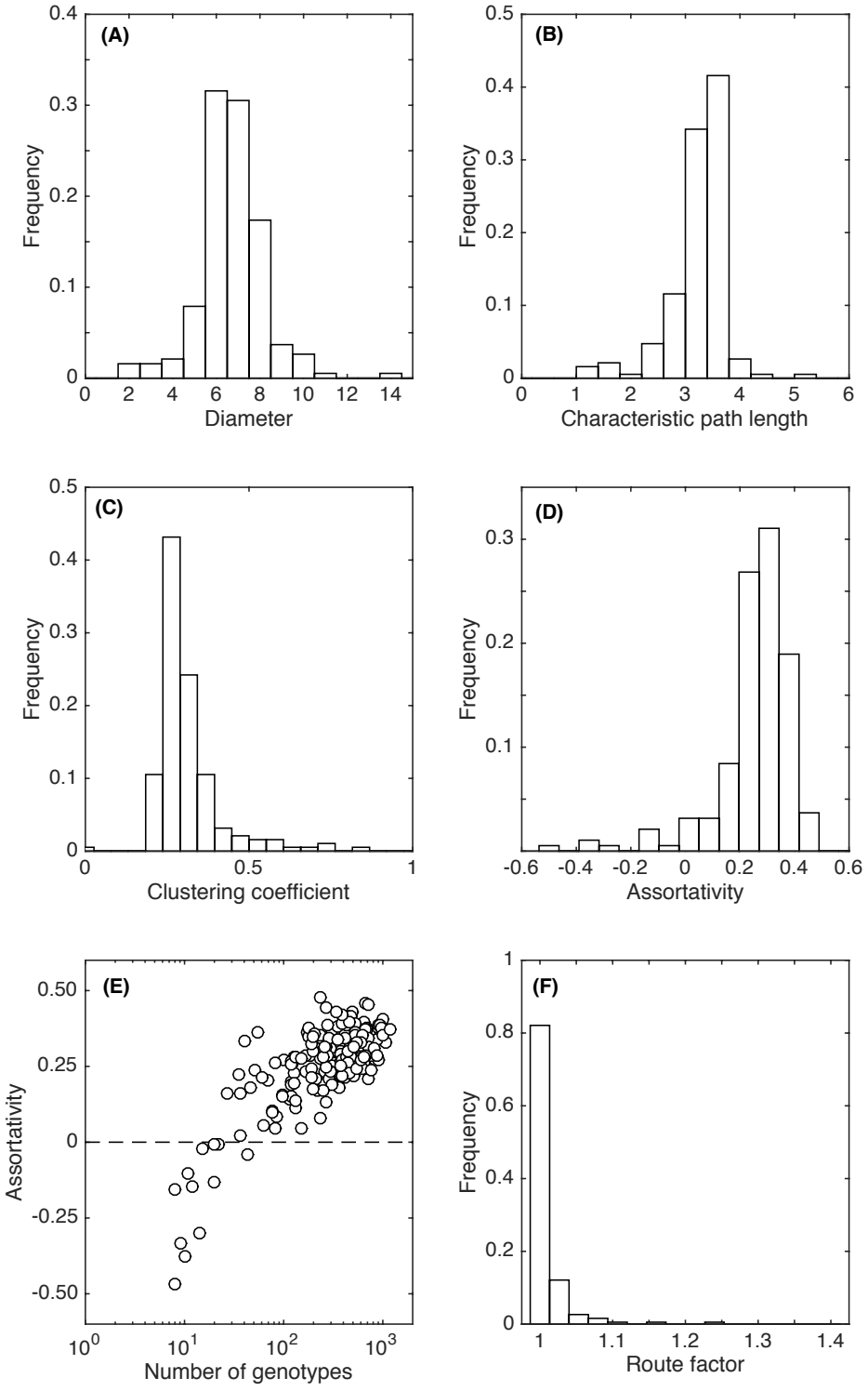

Figure 1. Intra-network statistics for 190 TFs from *M. musculus*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, and (D) assortativity. (E) Assortativity (vertical axis) and its relationship to the number of genotypes in the dominant genotype network (horizontal axis). The horizontal dashed line indicates an uncorrelated (non-assortative) mixing pattern. (F) The distribution of the genotype network route factor.

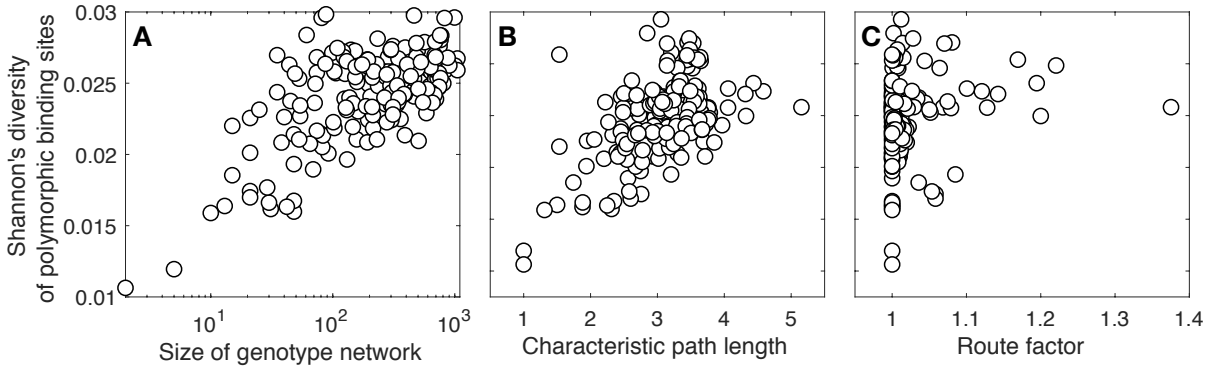

Figure 2. The structural properties of genotype networks are indicative of binding site diversity in extant populations of *A. thaliana***.** Shannon's diversity of a TF's polymorphic binding sites is shown in relation to (A) the number of nodes, (B) characteristic path length, and (C) route factor of its genotype network. The label of the y-axis applies to all panels.

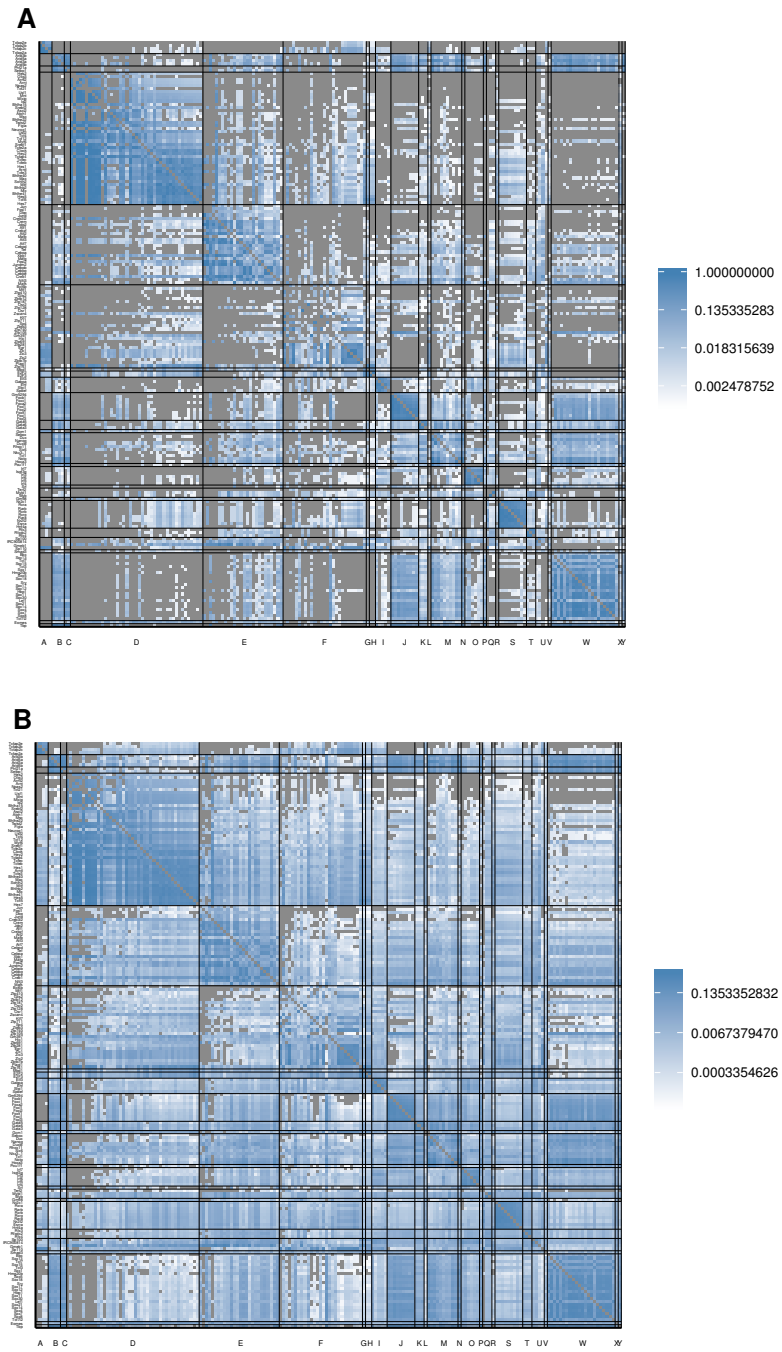

Figure 3. Matrices of inter-network relationships for the genotype networks of TF binding sites from *M. musculus*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. The rows and columns are grouped according to binding domain, which are ordered alphabetically on the horizontal axis: A, AP-2; B, ARID/BRIGHT; C, AT hook; D, bHLH; E, bZIP; F, C2H2 ZF; G, CxxC; H, E2F; I, Ets; J, Forkhead; K, GATA; L, GCM; M, Homeodomain; N, Homeodomain + POU; O, IRF; P, MADS box; Q, Myb/SANT; R, Ndt80/PhoG; S, Nuclear receptor; T, RFX; U, SAND; V, SMAD; W, Sox; X, T-Box; Y: TBP. Within the DNA-binding domain groups, the rows and columns are ordered by the size of each TF's dominant genotype network, such that network size increases from top to bottom and from left to right. Labels on the vertical axis indicate the name of the TFs, which can be read on the computer by zooming in. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

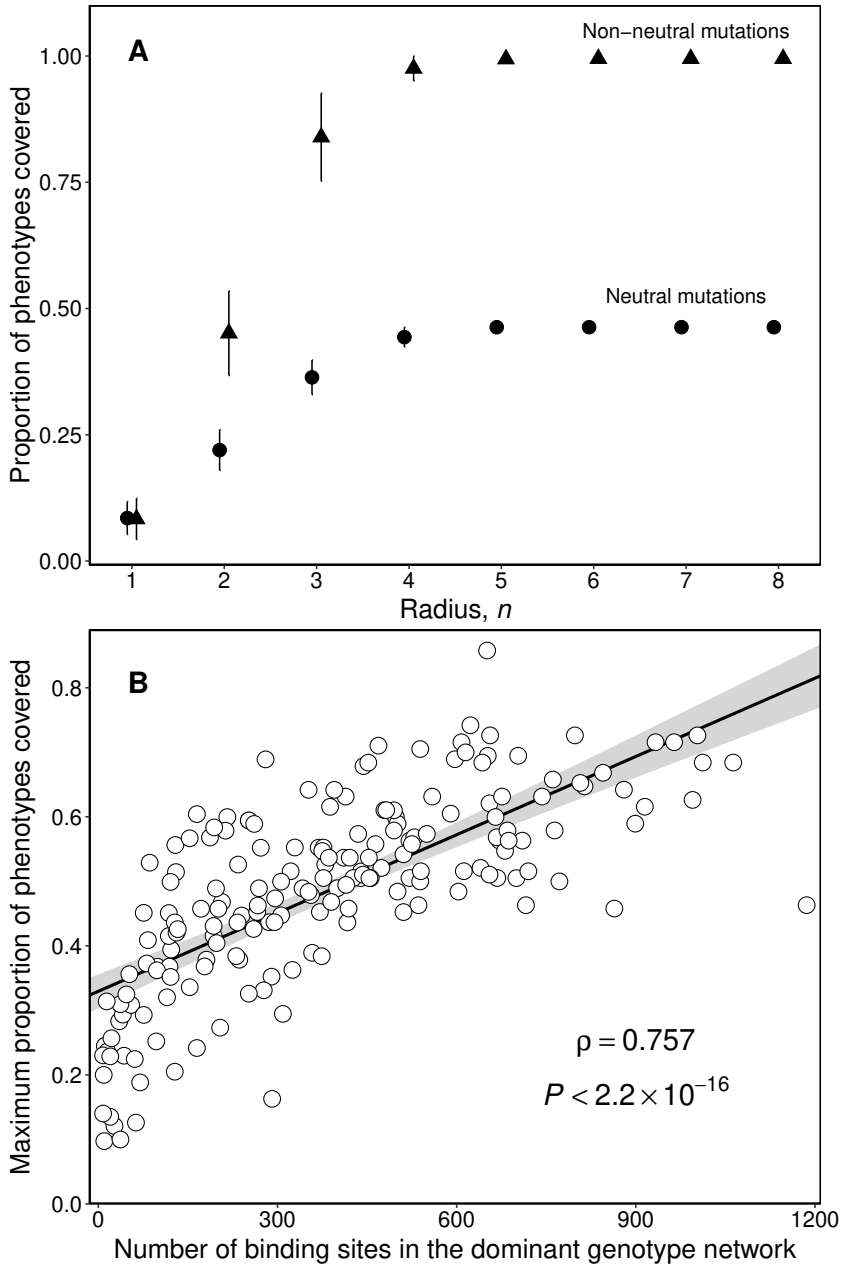

Figure 4. Phenotype space covering. (A) The proportion of phenotypes covered as a function of the mutational radius n from a given binding site, averaged across all binding sites of the murine TF Sp110. The maximum proportion of phenotypes covered plateaus at a much lower level when considering just neutral mutations than when considering non-neutral mutations. Error bars are the standard deviations of the mean. (B) The maximum proportion of phenotypes covered by neutral mutations as a function of the number of binding sites in the dominant genotype network, for all 190 murine TFs. The black line shows the fitted linear regression to the data ($R^2 = 0.516$) and the shaded grey area denotes 95% confidence intervals. The figure also shows the Spearman's correlation and its associated P-value.

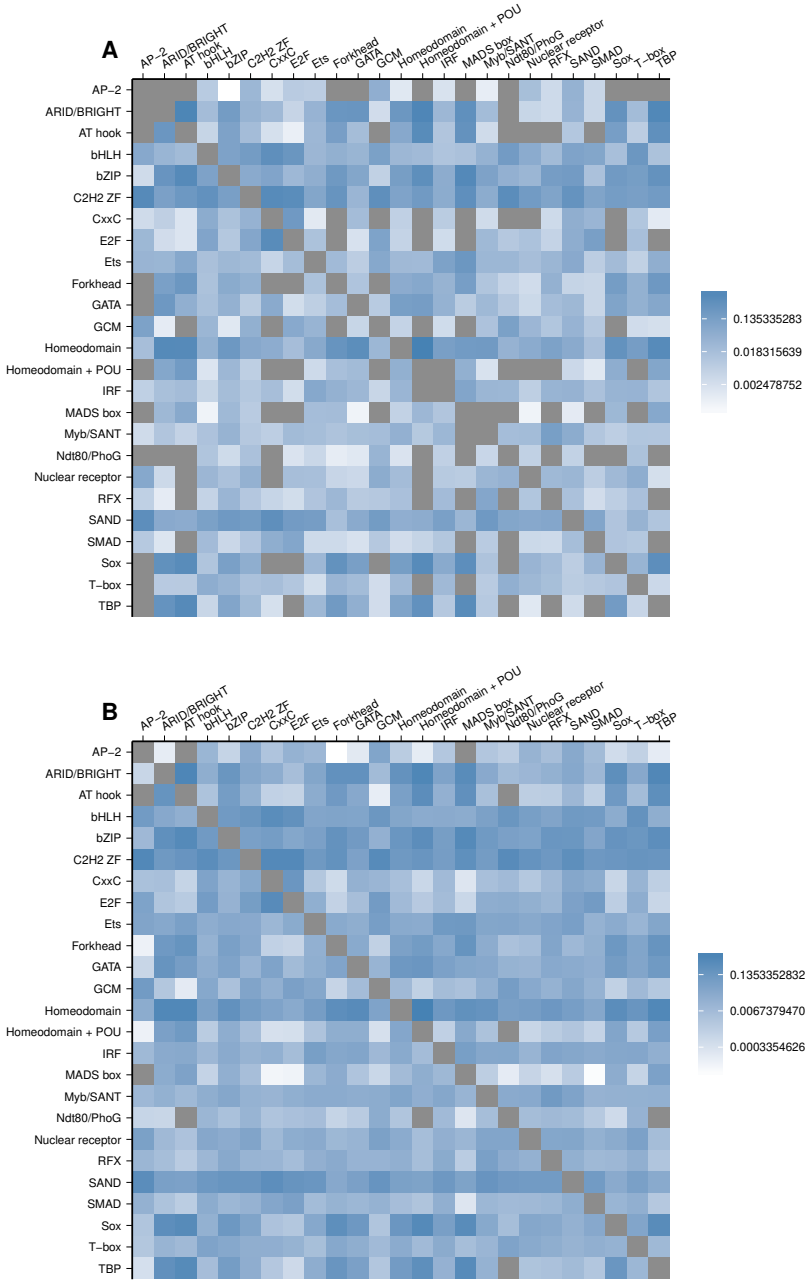

Figure 5. Matrices of inter-network relationships for the genotype networks of binding domains from *M. musculus*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. Each row and column represents a different genotype network. Domains are ordered alphabetically. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

Supplementary information

Supplementary methods

The stochastic block model for network partitioning

The stochastic block model (SBM) is a probabilisitic generative model for networks (Holland et al. 1983; Nowicki and Snijders 2001). Under the SBM, all nodes are assigned to one of k groups, and the probability of an edge connecting any pair of nodes depends only upon the nodes' group memberships. The pattern of edges can therefore be described by a single $k \times k$ "mixing matrix," in which each element p_{rs} gives the interaction probability between groups r and s (i.e., the probability that an edge exists between a node from group r and a node from group s.)

Using statistical inference (Peixoto 2014), we determined the maximum likelihood group assignment for each of the nodes in each genotype network. For a given assignment, the maximum likelihood interaction probability between groups r and s is given by the observed number of edges between the groups divided by the number of possible edges between the groups. That is,

$$p_{rs} = \frac{e_{rs}}{n_r n_s}, \tag{11}$$

where e_{rs} is the number of edges connecting nodes in group r to nodes in group s, and n_r and n_s are the number of nodes in groups r and s, respectively. Because we set the number of groups to $k = 2$, we have just three group interaction probabilities p_{11}, p_{12} and p_{22}, because the network is undirected. By comparing these probabilities, we can determine the type of structure the groups represent. For the two-group case there are three possibilities: $p_{11} > p_{12} < p_{22}$ (assortative), $p_{11} < p_{12} > p_{22}$ (disassortative), and $p_{11} > p_{12} > p_{22}$ (core-periphery) (Zhang et al. 2015).

Introduced in Peel et al. (2017), the block model entropy significance test provides a means for identifying whether node attributes are distributed randomly across a network. The test works by partitioning a network into groups of nodes that have the same node attribute value (for continuous-valued attributes, we form groups by discretizing the values into bins). Using this partition $\mathcal{M}$, we calculate the mixing matrix using Eq. (11). As a test statistic, we calculate the SBM entropy $\mathcal{H}$,

$$\mathcal{H}(\mathcal{M}) = -\left[\sum_{rs} e_{rs} \log p_{rs} + (n_r n_s - e_{rs} \log(1 - p_{rs})\right]. \tag{12}$$

High entropy indicates that node attributes are not correlated with network structure. Low entropy indicates that there is a correlation between the node attributes and the network structure. To determine if this correlation is statistically significant, we compare the observed entropy against a null distribution of entropy values. We obtain this distribution by randomly permuting node attributes, resulting in new partitions $\{\pi\}$ and corresponding mixing matrices. Importantly, this choice of null model preserves both the observed network structure and the relative frequencies of attribute values, but removes any correlation between the two. The result is a standard empirical p-value, defined as

$$p = Pr[H(\pi) \leq H(\mathcal{M})]. \tag{13}$$

Smaller p-values indicate a lower plausibility that a random permutation of the node attributes could describe the network structure as well as the observed distribution of node attributes.

Binding affinity partitions

We used the SBM partitions as a baseline for building node partitions that are based on binding affinities. For each genotype network, we attributed a categorical label to every node, indicating its SBM group. We chose "0" for nodes in the most assortative group and "1" for other nodes. This labeling also induces a partitioning of the binding affinities into two groups.

For each genotype network, we performed a logistic regression of the SBM partitioning of binding affinities. Using $b_{\min}$ and $b_{\max}$ to denote the minimum and maximum binding affinity values within a given genotype network, the regression resulted in a classifier $C : [b_{\min}, b_{\max}] \to [0, 1]$ that we trained on the empirical data. This classifier provided the likelihood that a given binding affinity value belonged to one SBM group or the other. In order to distinguish between "high" and "low" binding affinities, we chose the critical value b^* to be the binding affinity at which the classifier distinguished between groups, i.e. $C(b^*) = 0.5$. We used b^* in order to obtain a binding affinity partition, with nodes having binding affinities less than or equal to b^* in a group labeled g_{low}, and nodes with binding affinities greater than or equal to b^* in another group labeled g_{high}.

To test the statistical significance of binding affinity with respect to the structure of a genotype network, we again used the block model entropy significance test, such that in Eqs. 11 and 12, groups r and s were replaced with groups g_{low} and g_{high}.

Table S2. We show the number of genotype networks that have a partition that exhibits a particular group structure according to a partitioning method based on a stochastic block model.

Species	Group structure		
	Core-periphery	Assortative	Disassortative
A. thaliana	1 (0.46%)	213 (98.16%)	3 (1.38%)
N. crassa	1 (0.85%)	117 (99.15%)	0 (0.00%)
M. musculus	1 (0.53%)	186 (97.89%)	3 (1.58%)

Supplementary results

Genotype network partitions

Some TFs exhibit dual modes of binding specificity, such that they have a primary preference to one set of binding sites, as well as a secondary preference to a different set of binding sites (Badis et al. 2009). In some cases, these sets of sequences bind the TF with similar affinity, whereas in others the primary set binds the TF with higher affinity than the secondary set (Badis et al. 2009). These observations motivated us to perform an exploratory analysis of genotype network partitions — distinct groups of nodes that have more edges within them than between them (Zhang et al. 2015) — as these may reflect dual modes of binding specificity, or other nuances of TF-DNA interactions, such as DNA shape readout (Rohs et al. 2009).

To determine if partitions exist for genotype networks of TF binding sites, we took two approaches. In the first, we used a partitioning method that is based on a stochastic block model (Zhang et al. 2015). This method assigns each genotype in a genotype network to one of two groups (labeled g_1 and g_2), and uses a 2×2 "mixing matrix" to describe the structure of the network. This symmetric matrix contains the probabilities of observing edges between genotypes from the same group ($p_{g_1 g_1}$ and $p_{g_2 g_2}$) and between different groups ($p_{g_1 g_2}$). The method uses maximum likelihood to find the partition and mixing matrix that best explain the structure of the genotype network (Materials and Methods). The resulting probabilities of the mixing matrix can be used to classify each genotype network as exhibiting an assortative group structure ($p_{g_1 g_1} > p_{g_1 g_2} < p_{g_2 g_2}$), a disassortative group structure ($p_{g_1 g_1} < p_{g_1 g_2} > p_{g_2 g_2}$), or a core-periphery group structure ($p_{g_1 g_1} > p_{g_1 g_2} > p_{g_2 g_2}$) (Zhang et al. 2015). This approach is related to modularity optimization, and even produces the same results under certain conditions — specifically when the partitions are assortative and the edge densities of the groups are the same (Newman 2016). In contrast, the stochastic block model faciltates the discovery of a broader

range of significant partitions, including core-periphery and disassortative group structures, in addition to assortative group structures.

We find that the vast majority of genotype networks in the mouse dataset (97.9%) exhibit an assortative group structure (Table S2). Thus, not only are these networks globally assortative by degree ($r > 0$, Fig. 1C), they are also partitionable into two groups that each have more edges within them than between them. The same is true for the *A. thaliana* and *N. crassa* TFs, of which 98% and 99% exhibit an assortative group structure, respectively (Table S2).

We next asked whether similar trends in group structure exist if we manually partition each genotype network according to binding affinity, rather than relying on the maximum likelihood approach described above. Our motivation is that TFs with dual modes of binding specificity sometimes bind the sequences in the primary set with higher affinity than those in the secondary set (Badis et al. 2009). We used the structural partition of each genotype network into the two groups g_1 and g_2 to find an affinity threshold that best separates the binding affinities of these groups (Materials and Methods). We used this threshold to label the genotypes as belonging to a high-affinity group g_{high} or to a low-affinity group g_{low}. We then constructed a mixing matrix that contains the probabilities of observing edges within groups ($p_{g_{\text{low}},g_{\text{low}}}$ and $p_{g_{\text{high}},g_{\text{high}}}$) and between groups ($p_{g_{\text{low}},g_{\text{high}}}$), calculated directly from each genotype network. We used this mixing matrix to test the null hypothesis H_0 that binding affinity is distributed uniformly at random with respect to the structure of the genotype network (Materials and Methods) (Peel et al. 2017). Thus, rejection of H_0 indicates that the binding affinity partition provides meaningful information about genotype network structure. Table S3 shows that H_0 is almost always rejected. On the rare occasion that H_0 is accepted, the genotype network is small (≤ 72 nodes), which again likely indicates finite-size effects. Additionally, we find in *M. musculus* that 62.1% of the binding affinity partitions exhibit a core-periphery group structure ($p_{g_{\text{high}},g_{\text{high}}} > p_{g_{\text{high}},g_{\text{low}}} > p_{g_{\text{low}},g_{\text{low}}}$), while 34.7% exhibit an assortative group structure ($p_{g_{\text{high}},g_{\text{high}}} > p_{g_{\text{high}},g_{\text{low}}} < p_{g_{\text{low}},g_{\text{low}}}$). Similar results are obtained for the *A. thaliana* and *N. crassa* TFs (Table S3). In sum, genotype networks of TF binding sites can be partitioned in multiple meaningful ways, and the resulting group structure depends upon how the partition is defined. An assortative group structure is uncovered by a structure-based partition, whereas a core-periphery group structure can be uncovered by an affinity-based partition. Whether and how these partitions relate to dual modes of binding specificity, or to other facets of TF-DNA interactions, requires further investigation. It is our hope that by making these partitions publicly available, other researchers may use them to improve

Table S3. The number of genotype networks that have a binding affinity partition that exhibits a particular group structure. We also test the null hypothesis H_0 that binding affinity is random with respect to genotype network structure, rejecting H_0 if $p < 0.05$.

Species	Group structure			H_0
	Core-periphery	Assortative	Disassortative	rejections
A. thaliana	134 (61.75%)	79 (36.41%)	4 (1.84%)	209 (96.31%)
N. crassa	67 (56.78%)	49 (41.53%)	2 (1.69%)	113 (95.76%)
M. musculus	118 (62.11%)	66 (34.74%)	4 (2.11%)	181 (95.26%)

our understanding of TF binding specificity, perhaps via the development of more sophisticated binding models.

Some sequences have fewer than 32 neighbors in genotype space

Of the 32,896 sequences in genotype space, 1,312 have fewer than 32 neighbors (Fig. S1D). This occurs when two different mutations to a sequence yield the same mutated sequence, forcing the prioritization of one mutation over another. If the mutations are of different types (i.e., a point mutation and an indel), we always prioritize the point mutation over indels because laboratory evolution experiments indicate that they occur more frequently than indels (Cartwright 2009; Chen et al. 2009).

The 1,312 genotypes with fewer than 32 neighbors fall into the following five groups:

1. There are $4^4 = 256$ sequences that are identical to their reverse complements, and 252 of these have 16 neighbors. Due to the symmetry of these sequences, the number of possible point mutations in them is reduced from 24 to 12. The reason is that a point mutation in position $i \leq 4$ is equivalent to a point mutation to the Watson-Crick pair in position $8-i+1$, after taking the reverse complement. For example, consider the point mutation `A` $\rightarrow$ `C` in the 1st position of `ACGTACGT`. This yields the same sequence (`CCGTACGT`) as a point mutation `T` $\rightarrow$ `G` in the 8th position, after taking the reverse complement of the mutated sequence. The symmetry of these sequences also reduces the number of possible indels from 8 to 4. For example, an indel separates the sequence `ACGTACGT` from `CGTACGTA`, such that an alignment will leave the 1st position of the former sequence and the 8th position of the latter sequence unaligned. An indel also separates the sequence `ACGTACGT` from `TACGTACG`, such that an alignment will leave the 8th position of the former sequence and the 1st position of the latter sequence unaligned. Since the sequences `CGTACGTA` and `TACGTACG` are reverse complements of one another, it is not possible for `ACGTACGT` to have both of these mutational neighbors. In sum,

these 252 sequences only have $12 + 4 = 16$ neighbors.

2. Of the 256 sequences that are identical to their reverse complements, four have 15 neighbors: `AAAATTTT`, `CCCCGGGG`, `GGGGCCCC`, `TTTTAAAA`. The reasons are the same as for the other 252 sequences, except that the number of possible indels is further reduced to 3. To understand why, consider aligning the sequence `AAAATTTT` with `AAATTTTT`. This alignment could either include a point mutation in the 4th position, or an indel that leaves the 1st position of the former sequence and 8th position of the latter sequence unaligned. For this reason, these four sequences have 15 neighbors.

3. There are two sequences with 24 neighbors: `AAAAAAAA` and `CCCCCCCC`. They have 24 neighbors because we prioritize point mutations: We consider that any mutation that might be caused by an indel is more likely to be caused by a point mutation.

4. There are 46 sequences with 30 neighbors. 41 of these are of the form `AAAAAAAC`, `AAAAAACC`, `AAAAACCC`, ... ,`ACCCCCCC`, for which the number of possible indels is reduced from 8 to 6 because 2 indels are superseded by point mutations. For example, consider the sequence `AAAAAAAC`, which can be aligned to the sequences `AAAAAACC` and `AAAAAAAA` using either a point mutation or an indel. The remaining five sequences are `ACACTGTG`, `AGAGTCTC`, `ATATTATA`, `CTCTGAGA`, `GTGTCACA`. These sequences also have the number of indels reduced from 8 to 6, but for a more complicated reason. As an example, consider the sequence `ACACTGTG`, which is separated by a single point mutation from `ACACAGTG`. The reverse complement of `ACACAGTG` is `CACTGTGT`, which can be aligned to `ACACTGTG` with an indel, a mutation that is superseded by the point mutation from `ACACTGTG` to `ACACAGTG`.

5. There are 1008 sequences with 31 neighbors. These sequences have one indel that is superseded by a point mutation. For example, consider the sequence `AAAACTTT`. A point mutation `C` $\rightarrow$ `G` in the 5th position results in the sequence `AAAAGTTT`, whose reverse complement `AAACTTTT` can be aligned to `AAAACTTT` via an indel. This indel is therefore not included in the neighborhood of `AAAACTTT`, reducing the number of neighbors to 31.

Supplementary figures

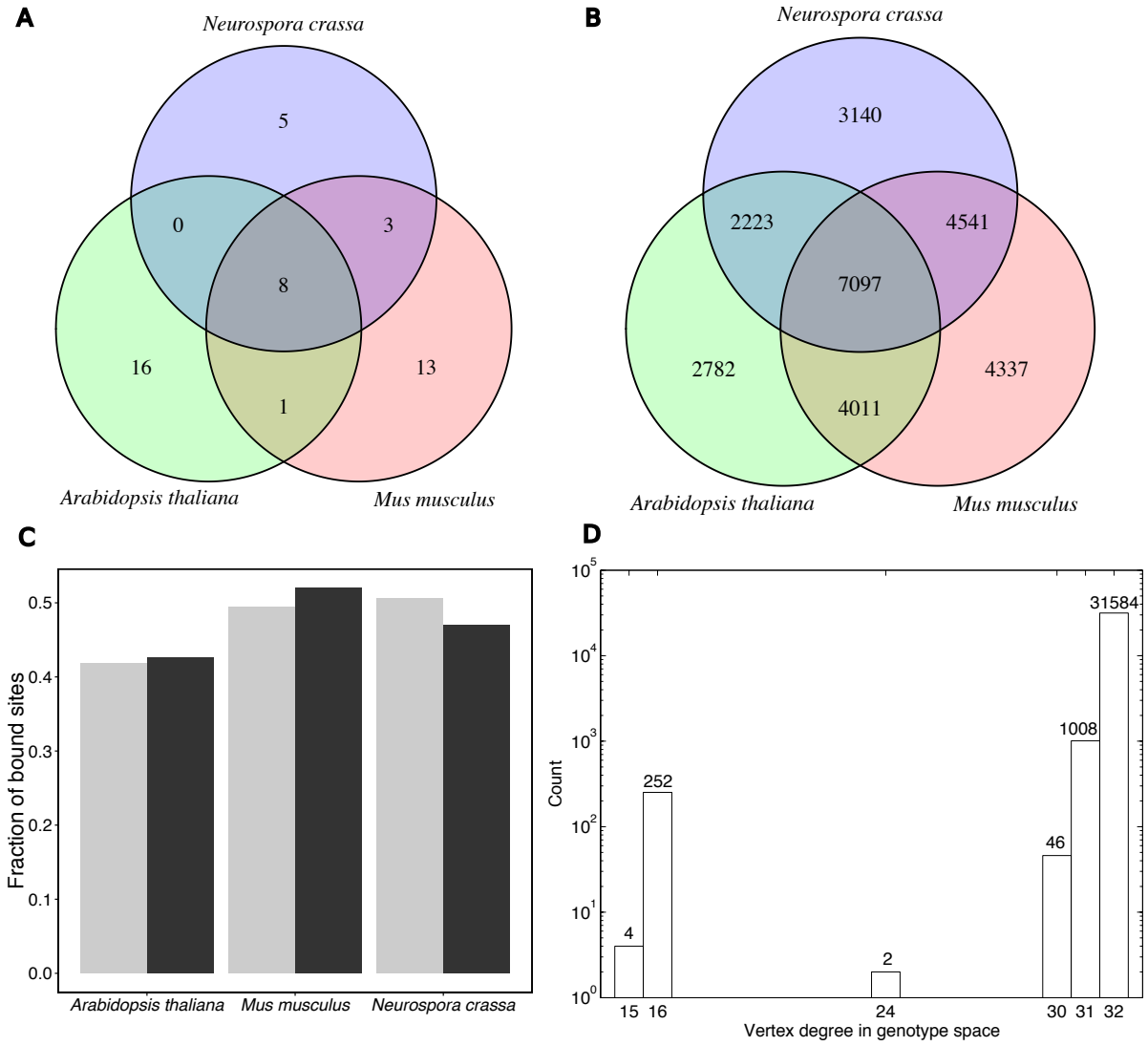

Figure S1. Data. (A) Venn diagram of the DNA-binding domains in the three species analyzed in this study (Table S1). (B) Venn diagram of the binding repertoires of the three species. (C) Amongst all of the sites that bind at least one TF in a given species, the gray bars show the fraction that bind TFs with binding domains that are unique to the species, and the black bars show the fraction that bind TFs with binding domains that are not unique to the species. Bar heights do not sum to one because there are sites bound by both types of TFs. (D) Genotype space is nearly regular. Bar plot of the degree distribution of Ω. Note the logarithmic scale of the y-axis and the counts above each bar.

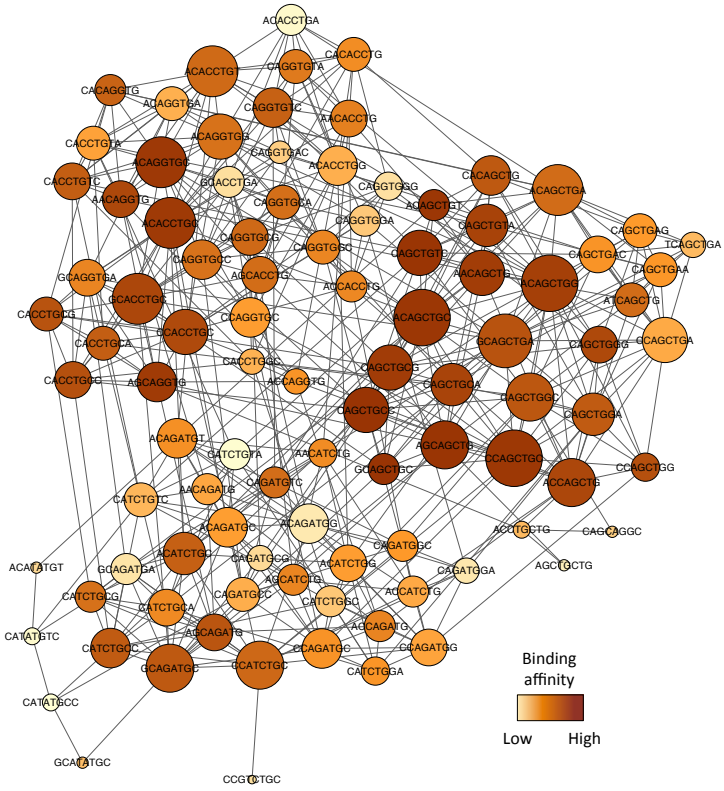

Figure S2. Genotype network of TF binding sites. (A) The dominant genotype network for the murine TF Ascl2. Each vertex corresponds to a DNA sequence that binds Ascl2 (E-score > 0.35). The color of a vertex indicates its binding affinity (darker = higher), while its size corresponds to the number of neighboring sequences (bigger = more). Two sequences are connected by an edge if they are separated by a single small mutation. This mutation may be a point mutation or an indel that shifts the entire binding site by a single position in either the $5'$ or $3'$ direction (Fig. S3).

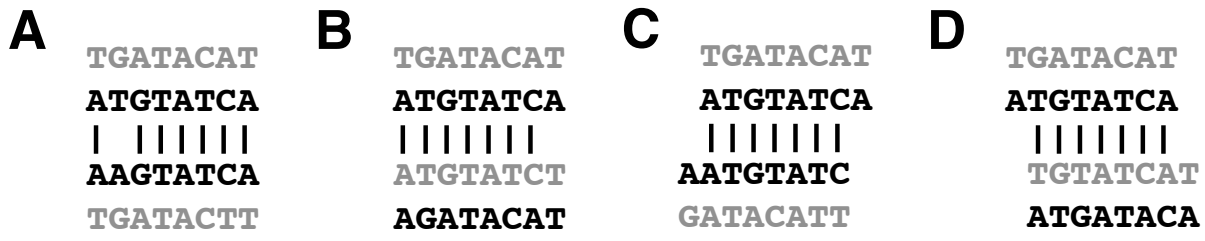

Figure S3. Two forms of mutation. We consider (A,B) point mutations and (C,D) indels that shift an entire, contiguous binding site by a single base. These mutations are illustrated by aligning four different sequences with `ATGTATCA` (top bold-font sequence in each panel). Since every sequence is merged with its reverse complement (gray font), the $4^8 = 65,536$ possible sequences of length eight can be represented by a library of only 32,896 sequences. Sequences that are members of this library are represented in bold font, while their reverse complements are represented in gray font.

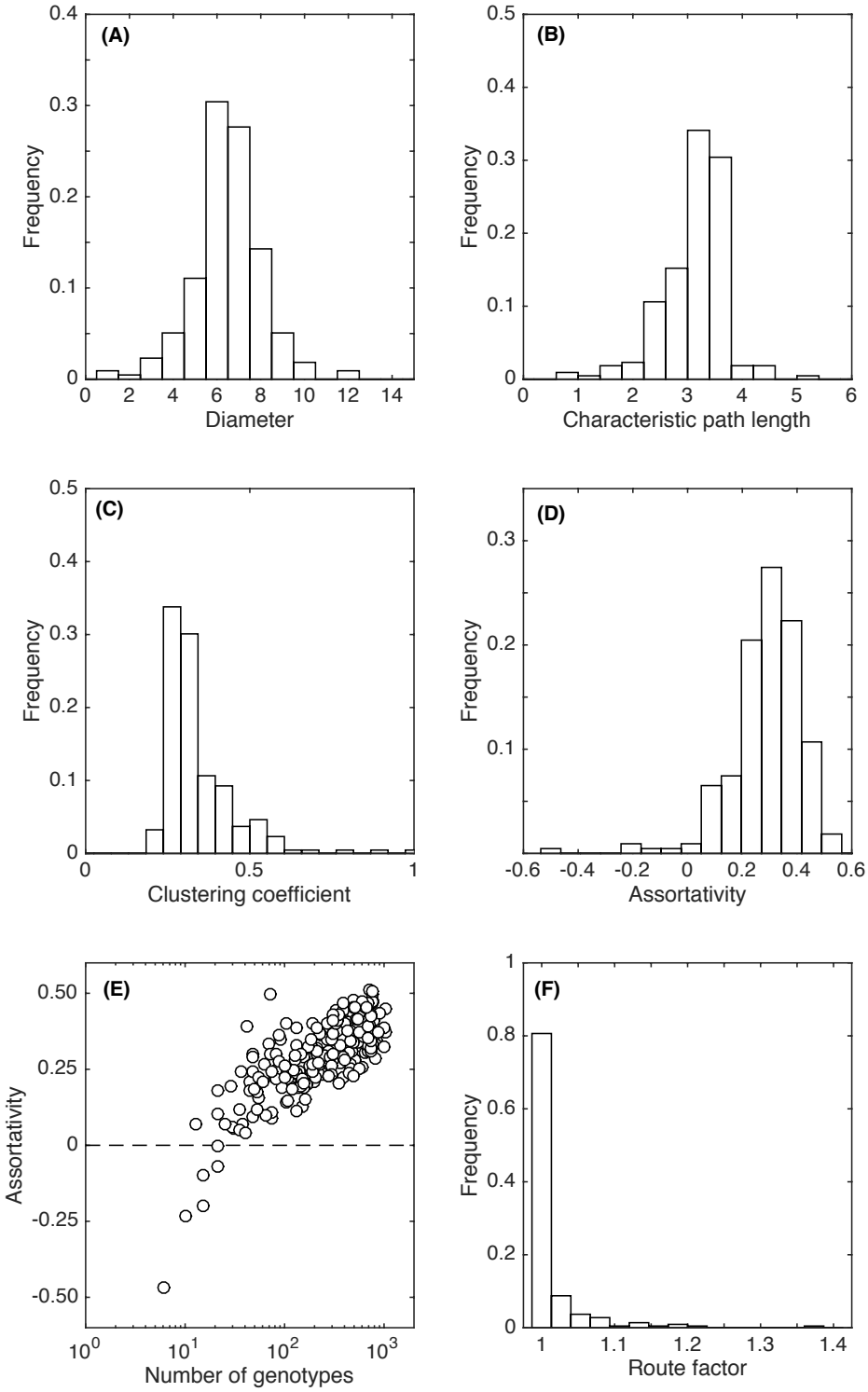

Figure S4. Intranetwork statistics for 217 TFs from *A. thaliana*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, and (D) assortativity. (E) Assortativity (vertical axis) and its relationship to the number of genotypes in the dominant genotype network (horizontal axis). The horizontal dashed line indicates an uncorrelated (non-assortative) mixing pattern. (F) The distribution of the genotype network route factor.

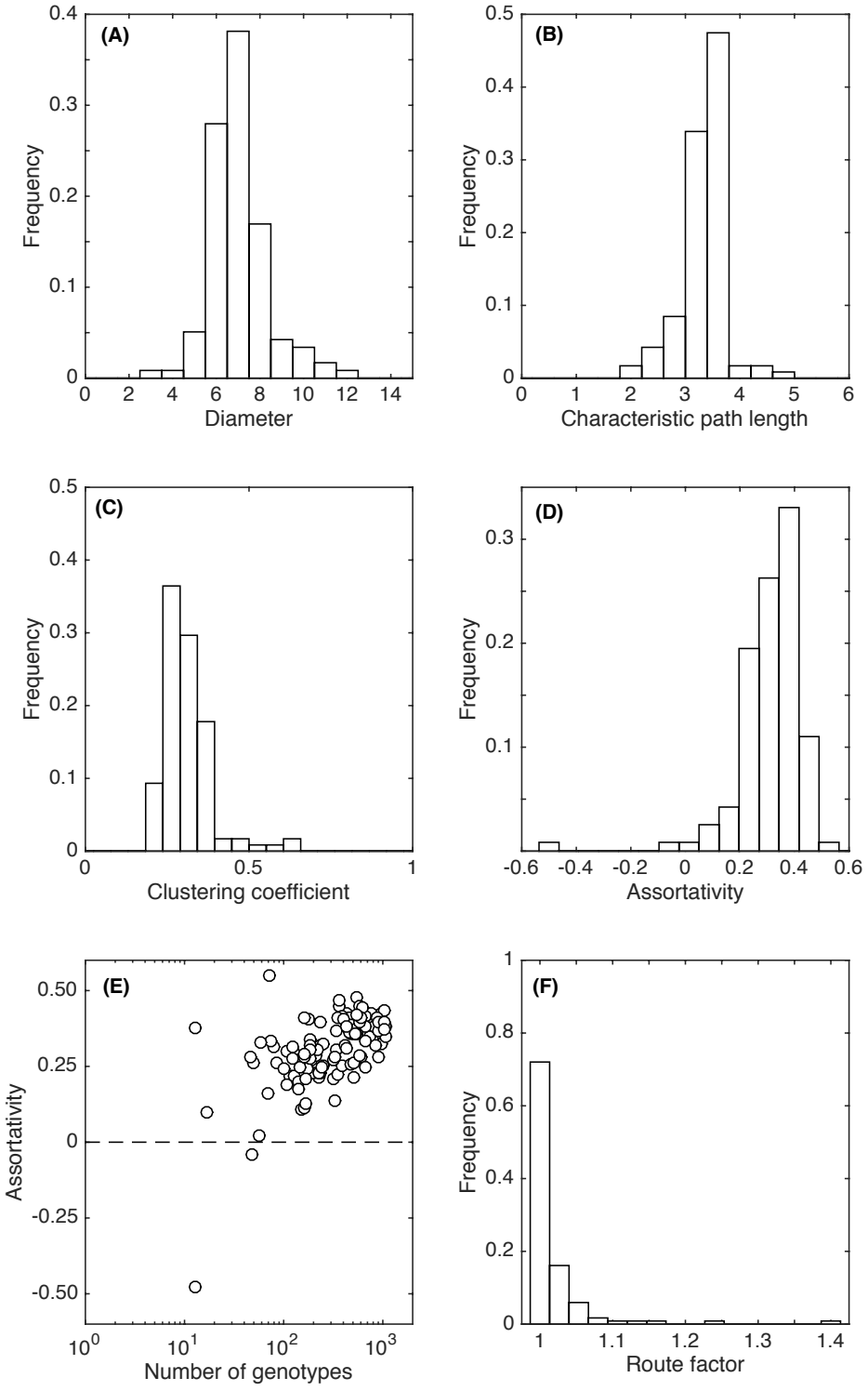

Figure S5. Intranetwork statistics for 118 TFs from *N. crassa*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, and (D) assortativity. (E) Assortativity (horizontal axis) and its relationship to the number of genotypes in the dominant genotype network (vertical axis). The horizontal dashed line indicates an uncorrelated (non-assortative) mixing pattern. (F) The distribution of the genotype network route factor.

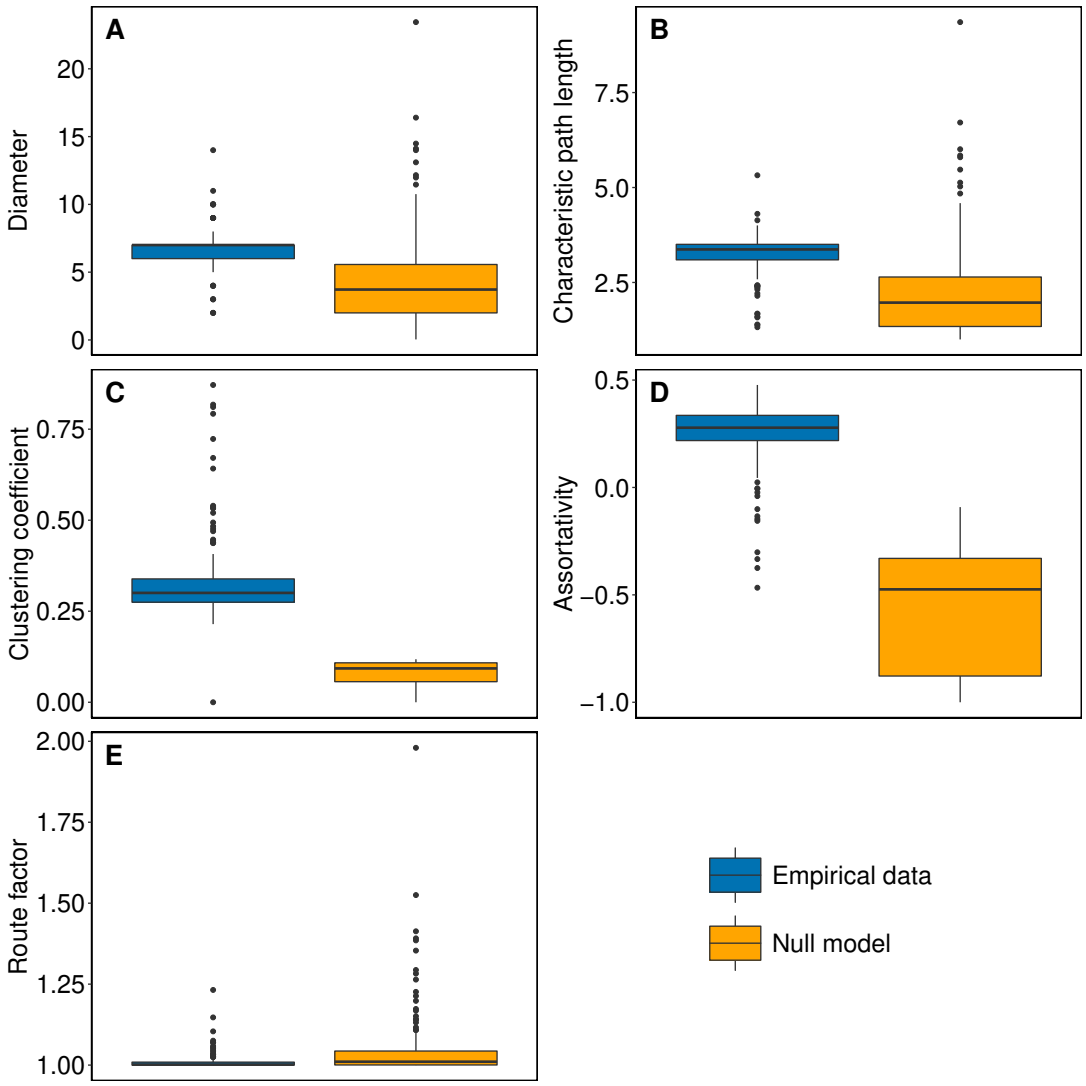

Figure S6. Comparison of intranetwork statistics to those of a null model for 190 TFs from *M. musculus*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, D) assortativity, and (E) route factor.

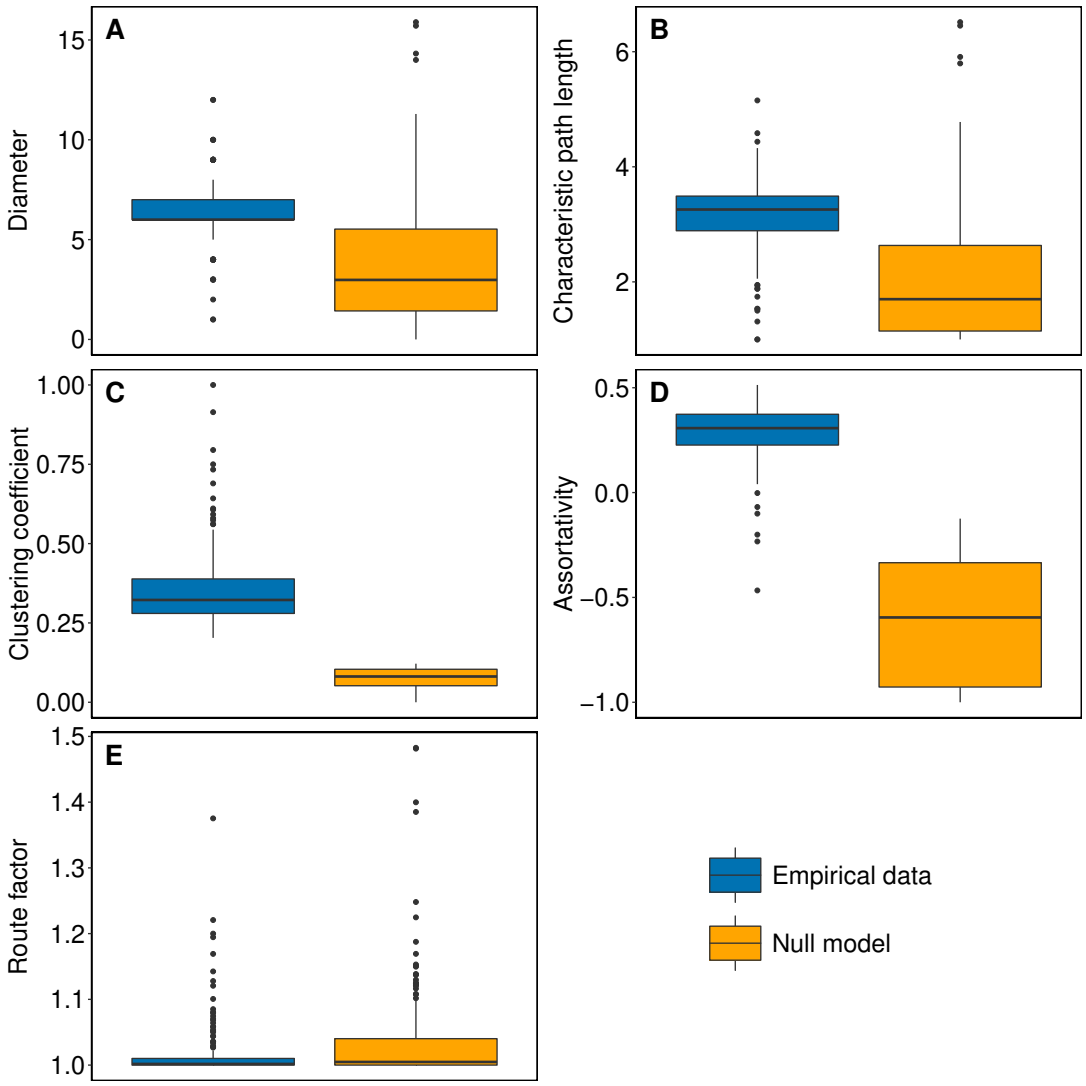

Figure S7. Comparison of intranetwork statistics to those of a null model for 217 TFs from *A. thaliana*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, assortativity, and (E) route factor.

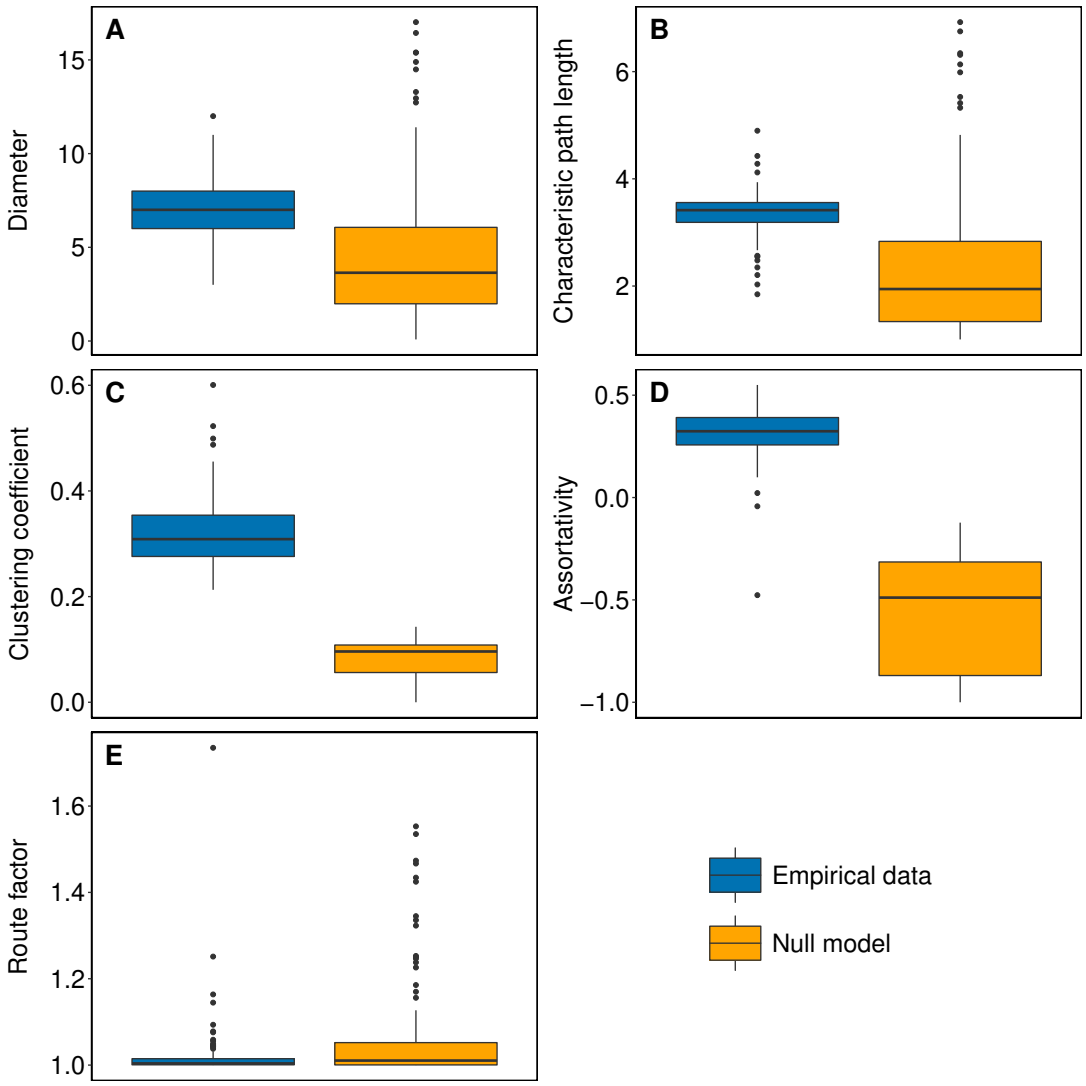

Figure S8. Comparison of intranetwork statistics to those of a null model for 118 TFs from *N. crassa*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, D) assortativity, and (E) route factor.

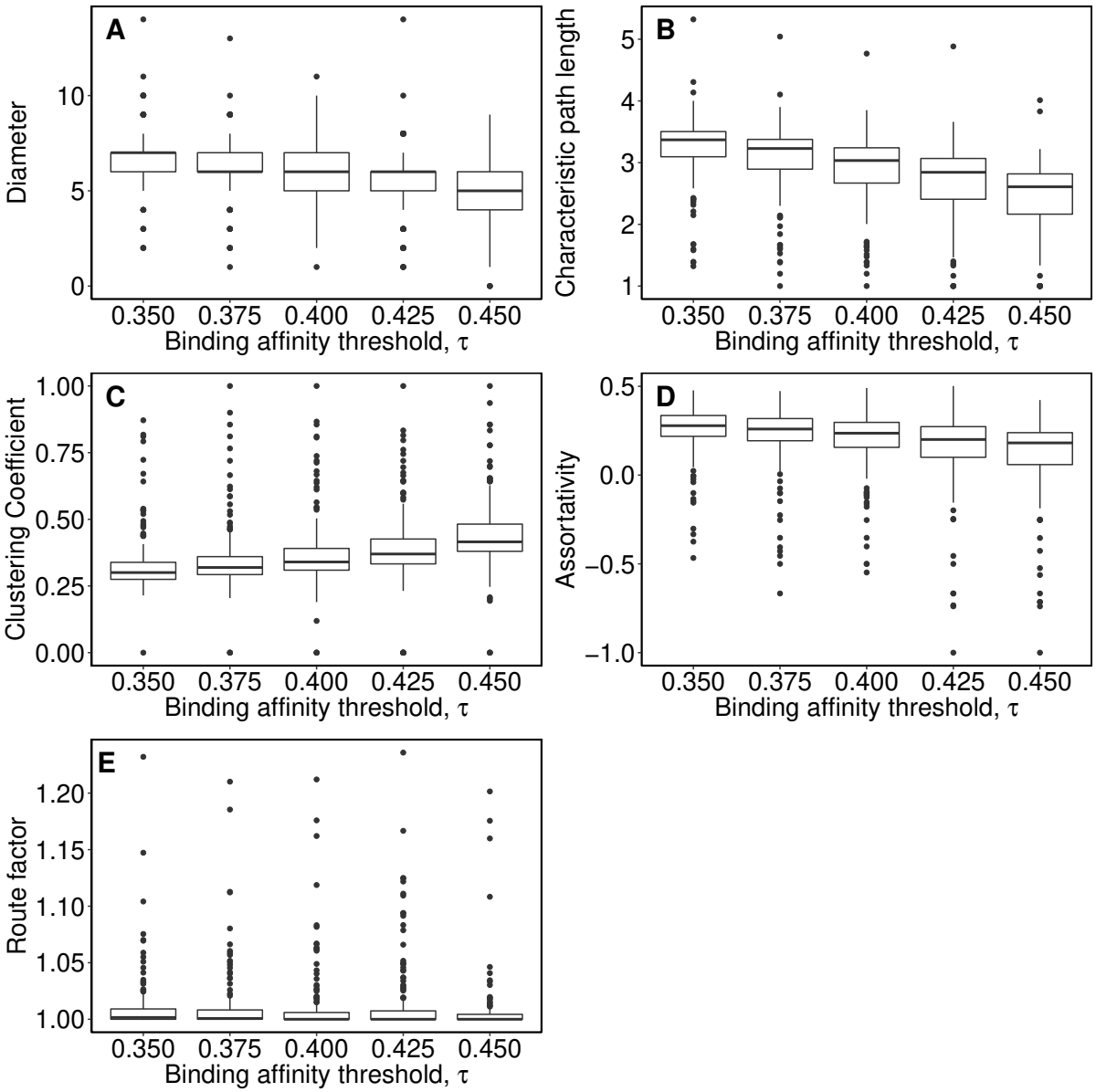

Figure S9. Intranetwork statistics with different binding affinity thresholds for 190 TFs from *M. musculus*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, D) assortativity, and (E) route factor, as a function of the binding affinity threshold. These data represent a sensitivity analysis of the results presented in Fig. 1.

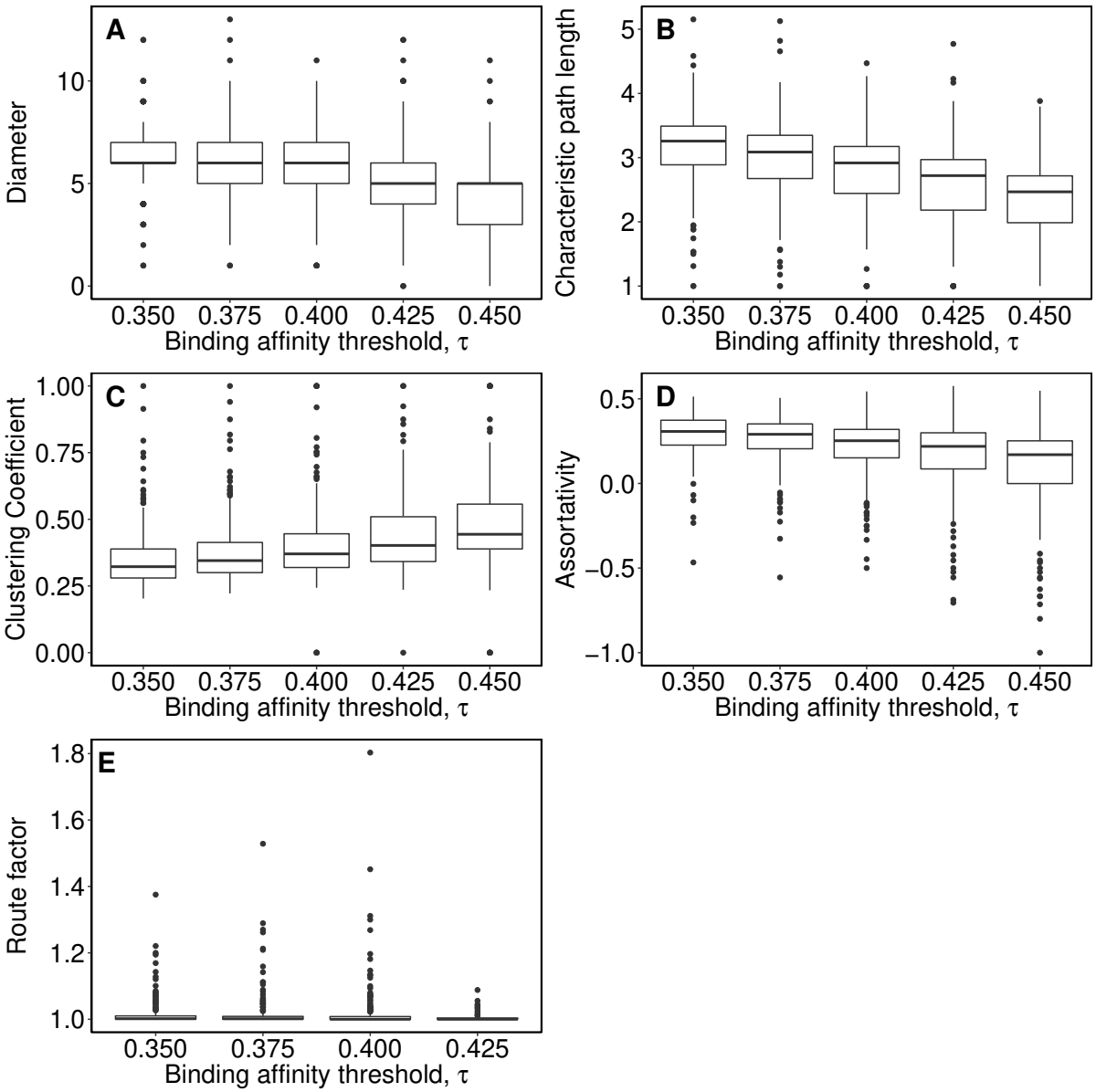

Figure S10. Intranetwork statistics with different binding affinity thresholds for 217 TFs from *A. thaliana*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, D) assortativity, and (E) route factor, as a function of the binding affinity threshold. These data represent a sensitivity analysis of the results presented in Fig. S4.

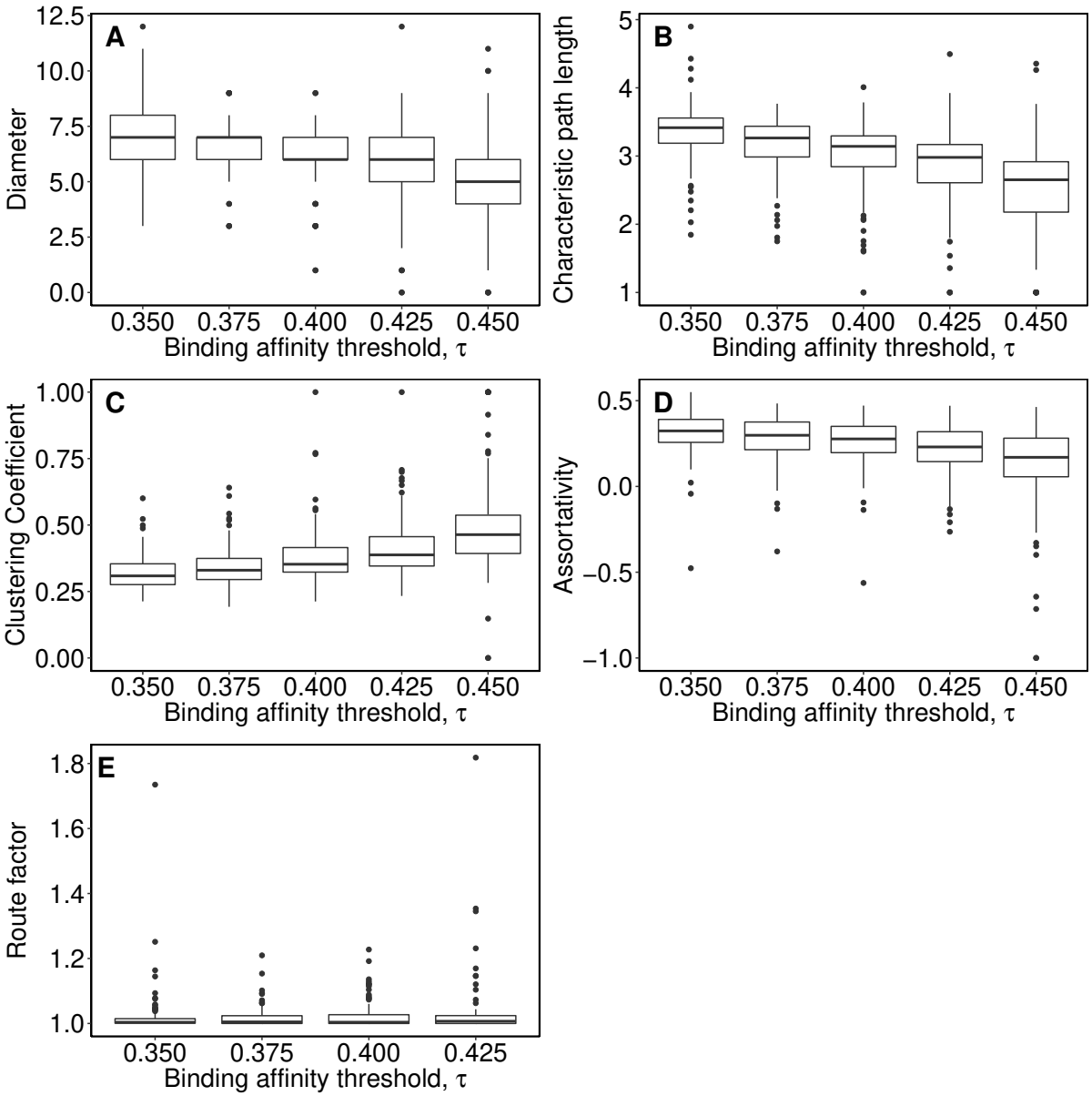

Figure S11. Intranetwork statistics with different binding affinity thresholds for 118 TFs from *N. crassa*. The distributions of genotype network (A) diameter, (B) characteristic path length, (C) clustering coefficient, D) assortativity, and (E) route factor, as a function of the binding affinity threshold. These data represent a sensitivity analysis of the results presented in Fig. S5.

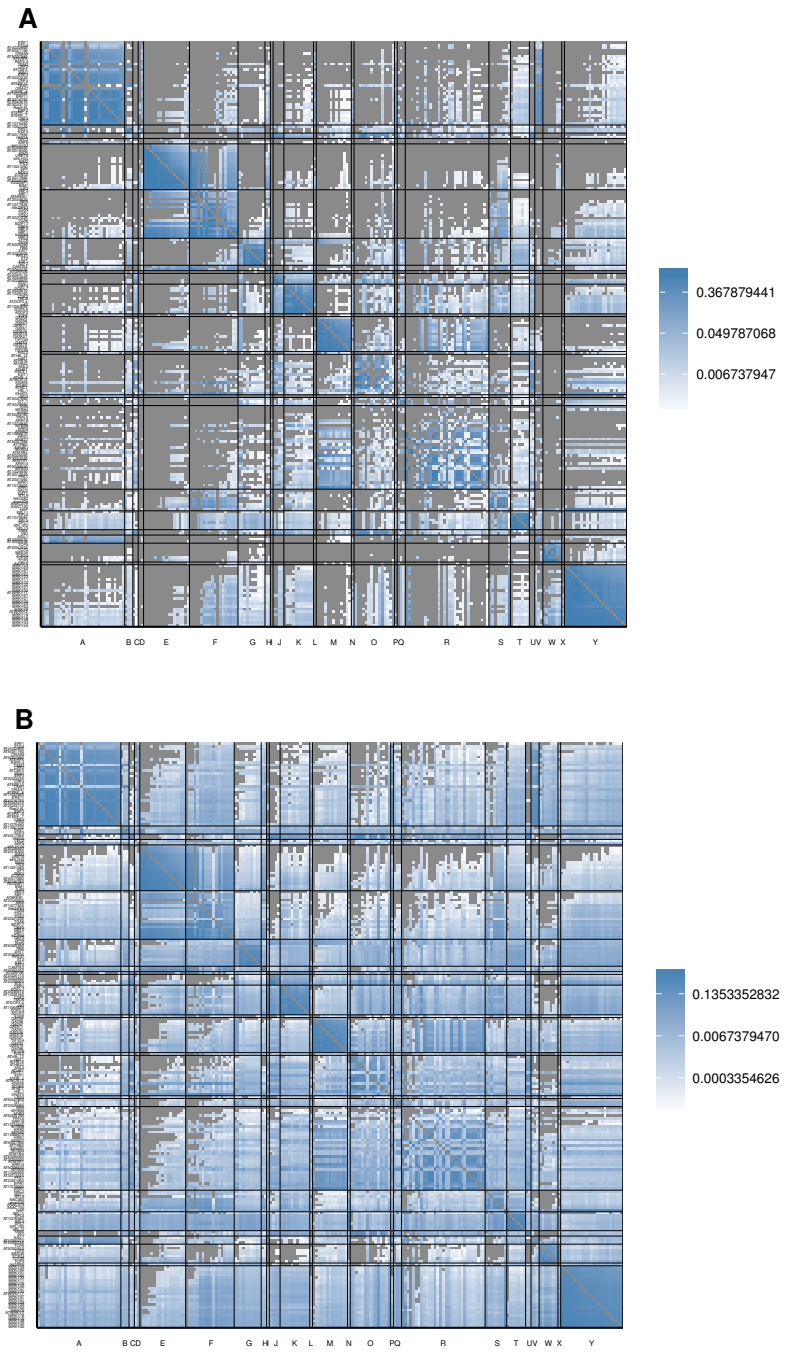

Figure S12. Matrices of internetwork relationships for the genotype networks of TF binding sites from *A. thaliana*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. The rows and columns are grouped according to binding domain, which are ordered alphabetically on the horizontal axis: A, AP2; B, AP2B3; C, AT hook; D, B3; E, bHLH; F, bZIP; G, C2H2 ZF; H, CG-1; I, CSD; J, CxC; K, Dof; L, E2F; M, GATA; N: GRAS; O, Homeodomain; P, LOB; Q, MADF; R, Myb/SANT; S, NAC/NAM; T, SBP; U, Sox; V, Storekeeper; W, TCP; X, WRC; Y, WRKY. Within the DNA-binding domain groups, the rows and columns are ordered by the size of each TF's dominant genotype network, such that network size increases from top to bottom and from left to right. Labels on the vertical axis indicate the name of the TFs, which can be read on the computer by zooming in. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

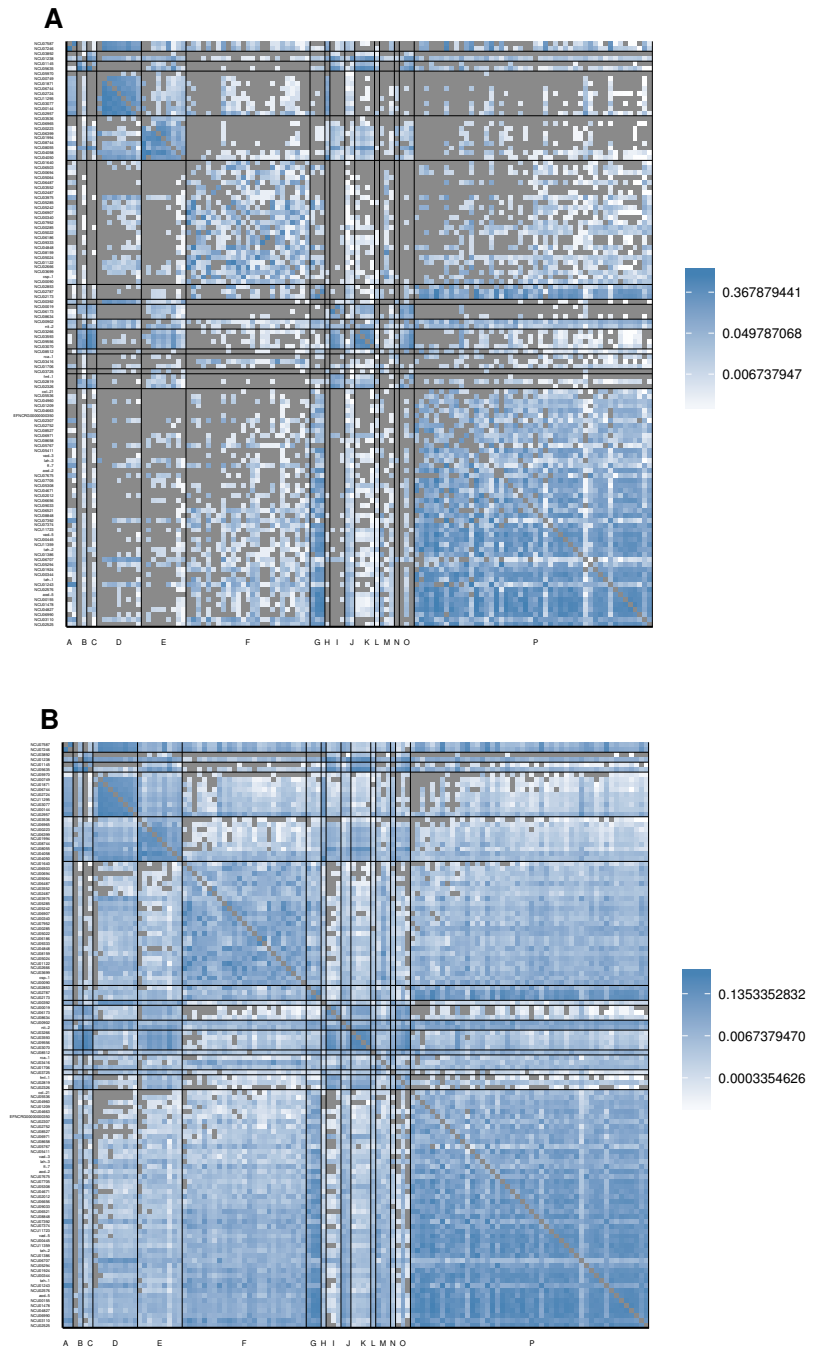

Figure S13. Matrices of internetwork relationships for the genotype networks of TF binding sites from *N. crassa*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. The rows and columns are grouped according to binding domain, which are ordered alphabetically on the horizontal axis: A, APSES; B, ARID/BRIGHT; C, AT hook; D, bHLH; E, bZIP; F, C2H2 ZF; G, C2H2 ZF + Zinc cluster; H, CENPB; I, Forkhead; J, GATA; K, Homeodomain; L, HSF; M, Myb/SANT; N, Ndt80/PhoG; O, Sox; P, Zinc cluster. Within the DNA-binding domain groups, the rows and columns are ordered by the size of the dominant genotype net=work, such that network size increases from top to bottom and from left to right. Labels on the vertical axis indicate the name of the TFs, which can be read on the computer by zooming in. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

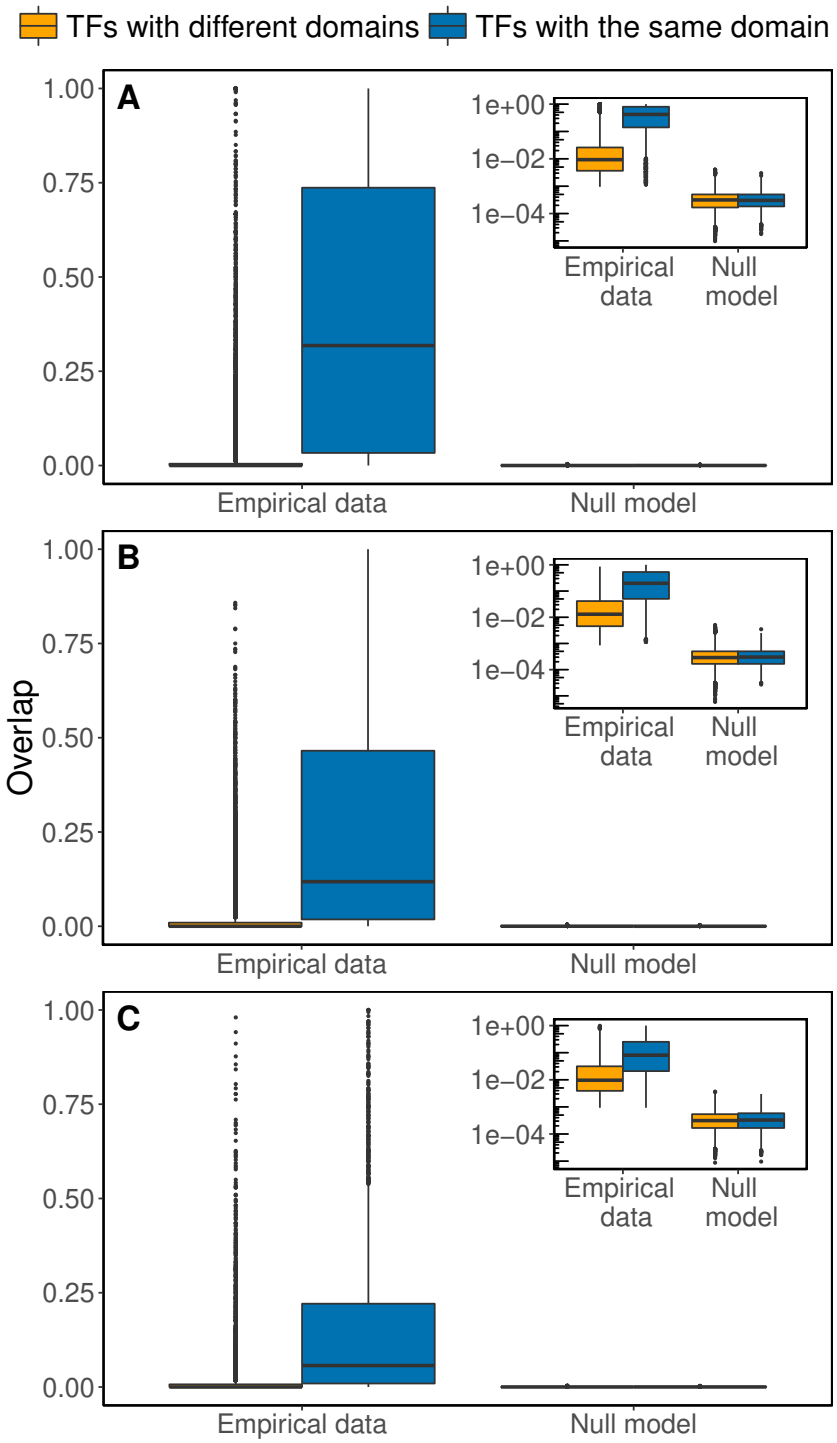

Figure S14. Comparison of overlap in the empirical data to that in the null model. The empirical data is the same as presented in Figs. 3A, S12A and S13A. TFs that belong to the same DNA-binding domain family show higher overlap than TFs belonging to different domain families, but this difference is not observed in the null model. This observation is made in all three species: (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*. Insets show the same data as the main panels, but with a logarithmically-scaled y-axis.

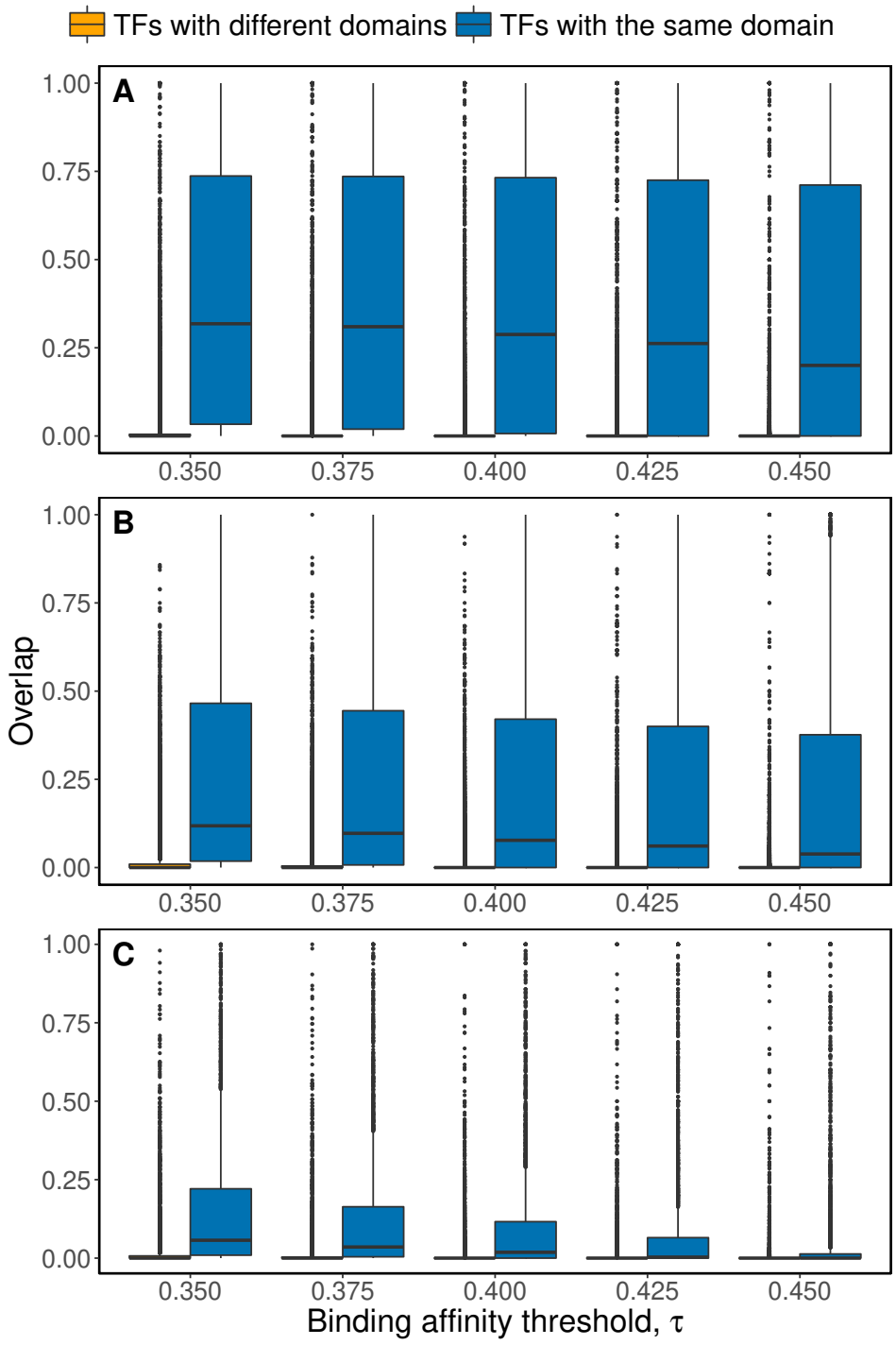

Figure S15. Genotype network overlap in relation to the binding affinity threshold. These data represent a sensitivity analysis of the results presented in Figs. 3A, S12A and S13A. TFs that belong to the same DNA-binding domain family show higher overlap than TFs belonging to different domain families. This observation does not change with binding affinity threshold in (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*.

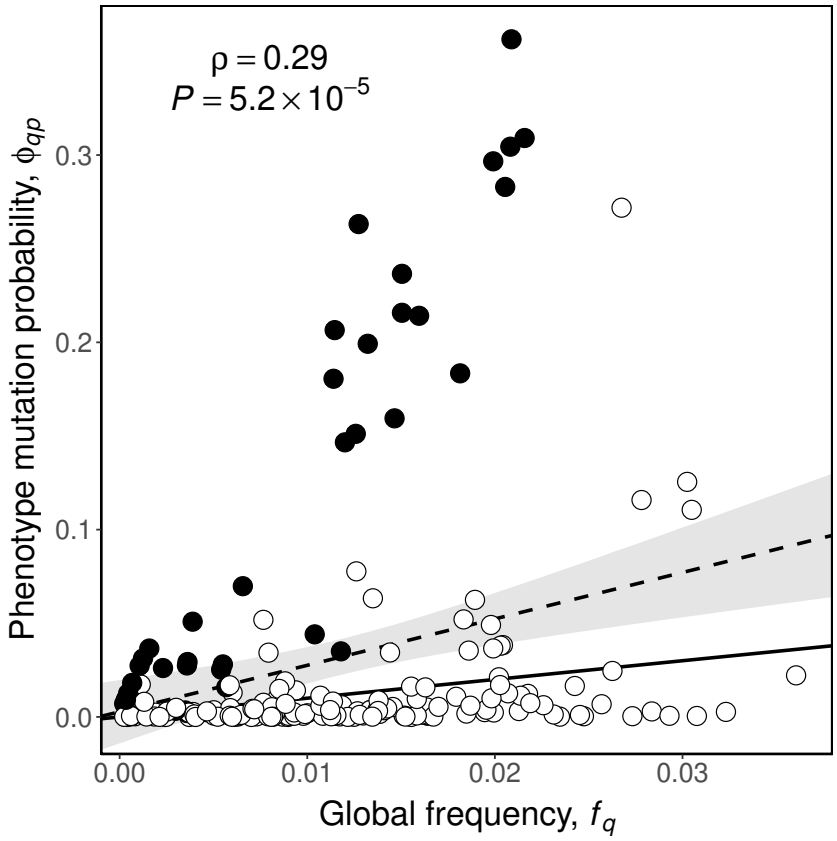

Figure S16. A simple null model does not provide a reasonable approximation to ϕ_{qp}. The probability ϕ_{qp} that a mutation to a genotype with phenotype p creates a genotype with phenotype q is shown in relation to the frequency f_q of phenotype q. The black line shows the null expectation that $\phi_{qp} = f_q$ (Greenbury et al. 2016). The dashed line shows the fitted linear regression to the data ($R^2 = 0.082$) and the shaded grey band around the line denotes 95% confidence intervals. The figure also shows the Spearman's correlation (ρ) between phenotype mutation probability and phenotype frequency, and its associated P value. Each circle represents the ϕ_{qp} of a different phenotype p, where phenotype q is always that of the murine TF Hes7. Black circles correspond to TFs with bHLH binding domain, and white circles correspond to TFs with a different binding domain. Half circles at the bottom of the panel denote pairs of phenotypes with $\phi_{qp} = 0$.

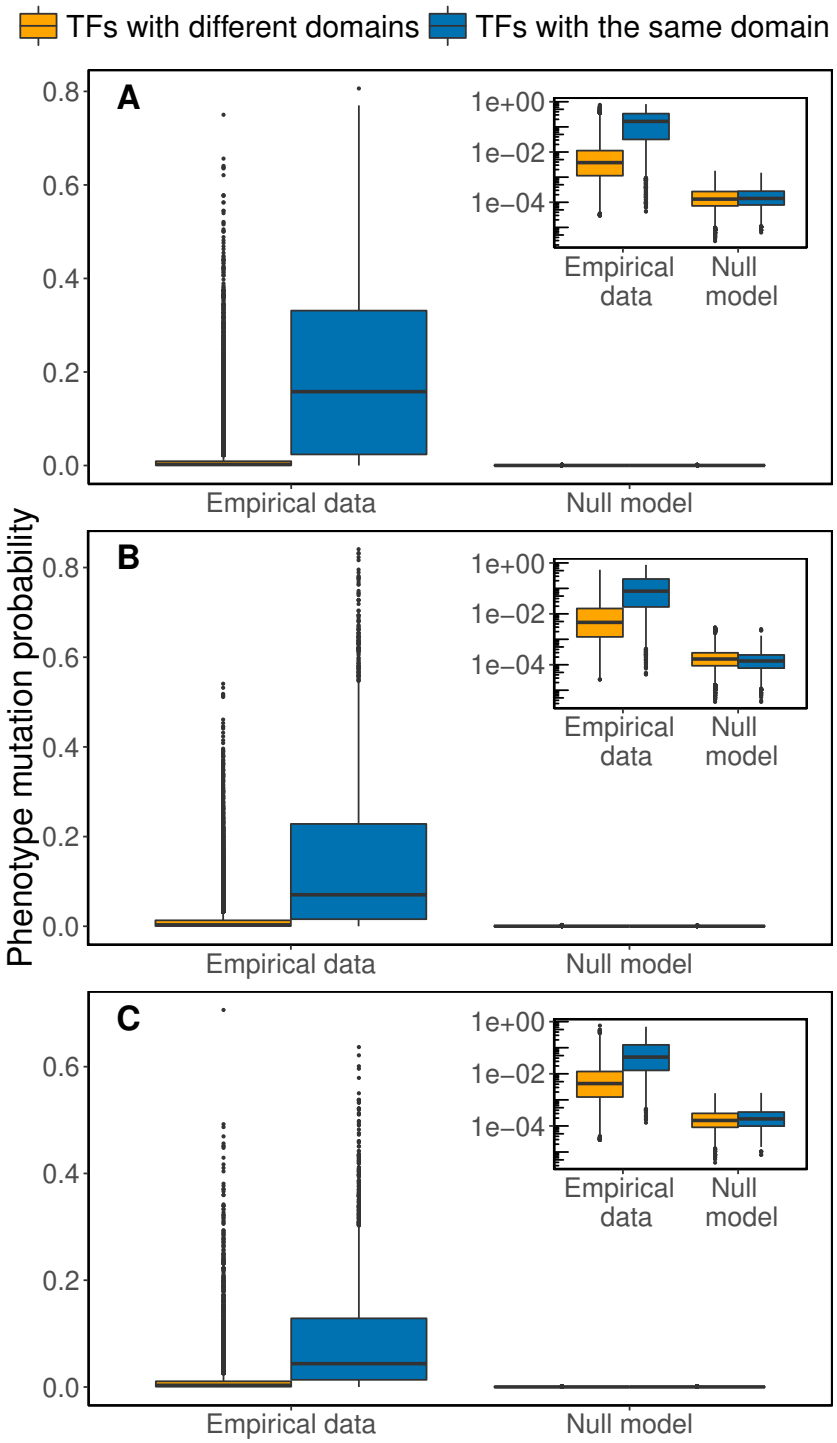

Figure S17. Comparison of phenotype mutation probabilities in the empirical data to those of a null model. The empirical data is the same as presented in Fig. 3B, S12B and S13B. TFs that belong to the same DNA-binding domain family show higher phenotype mutation probabilities than TFs belonging to different domain families, but this difference is not observed in the null model. This observation is made in all three species: (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*. Insets show the same data as the main panels, but with a logarithmically-scaled y-axis.

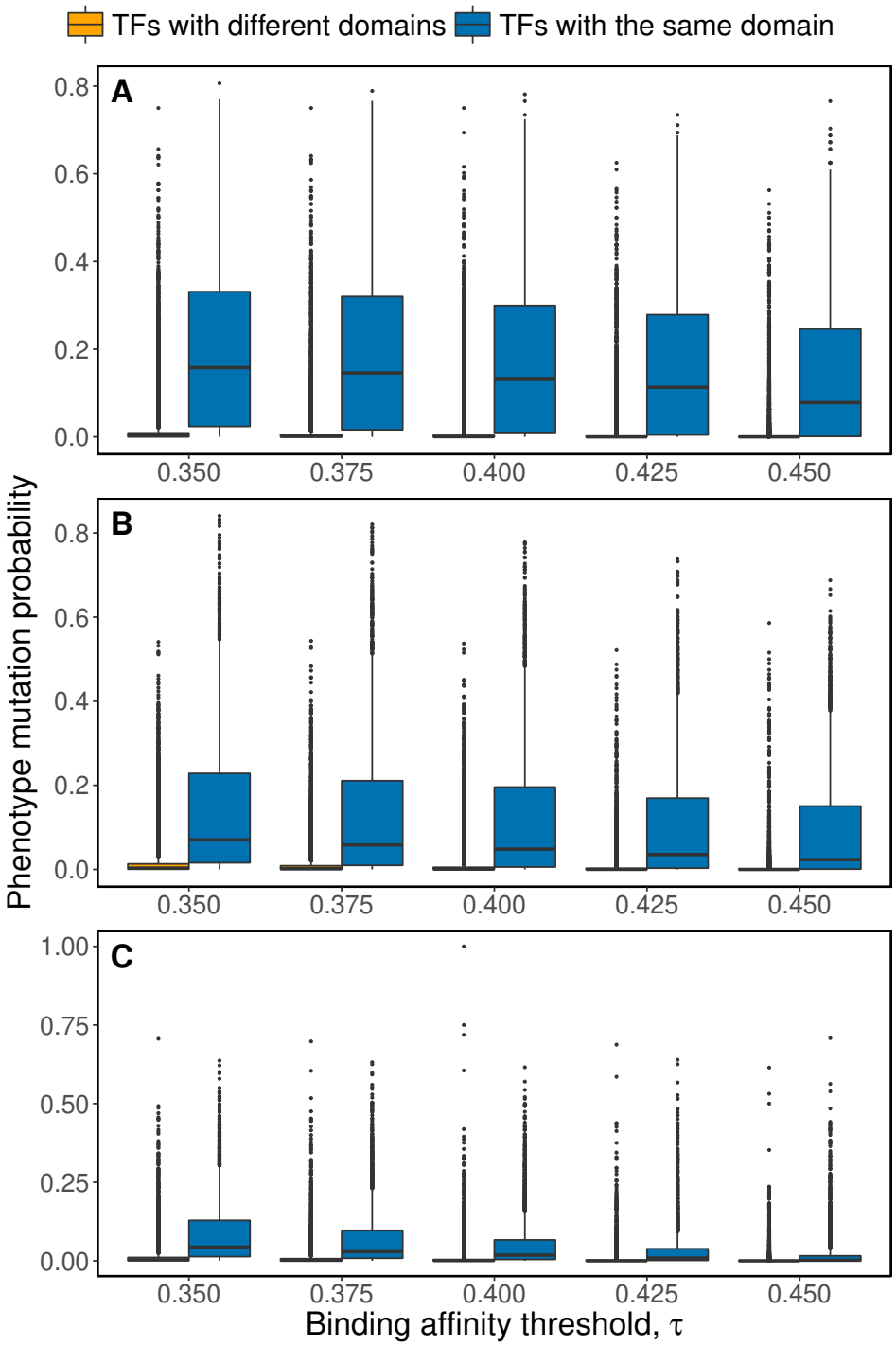

Figure S18. Phenotype mutation probabilities in relation to binding affinity threshold. These data represent a sensitivity analysis of the results presented in Fig. 3B, S12B and S13B. TFs that belong to the same DNA-binding domain family show higher phenotype mutation probability than TFs belonging to different domain families. This observation does not change with binding affinity threshold in (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*.

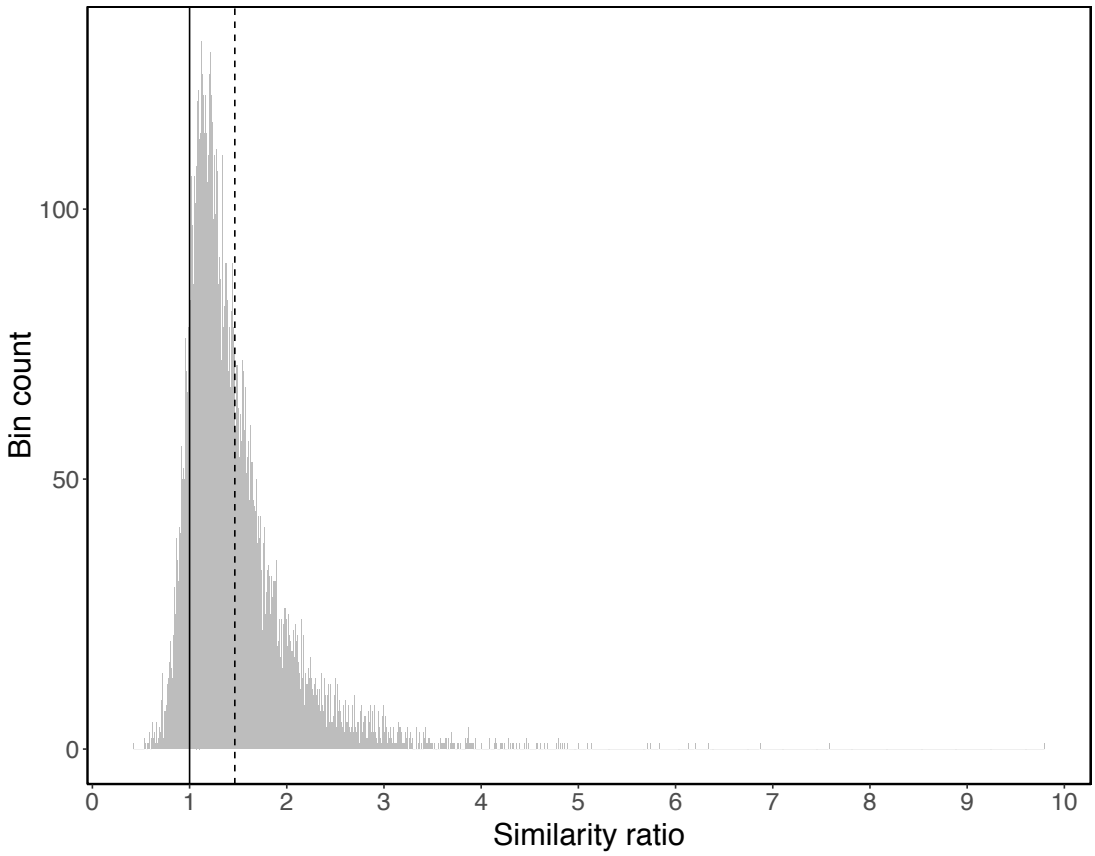

Figure S19. In *M. musculus*, the phenotypes found in the mutational neighborhoods of neutral neighbors are more similar than those of neutral pairs that are not neighbors. The distribution of the similarity ratio (Eq. 9) of the phenotype probability distributions (Eq. 6) is shown for neutral neighbors (n_1 and n_2) and neutral pairs that are not neighbors (n_1 and n_3). For this analysis, we considered all 9,207 pairs of neutral neighbors in the genotype network for Sp110, and sampled the same number of neutral pairs that are not neighbors. The mutational neighborhoods of n_1 and n_2 are more similar than those of n_1 and n_3, because the mean of the ratio (vertical dashed line) is larger than unity (vertical solid line). The standard error of this mean (0.006) is minute compared to the difference between the mean and unity (0.465).

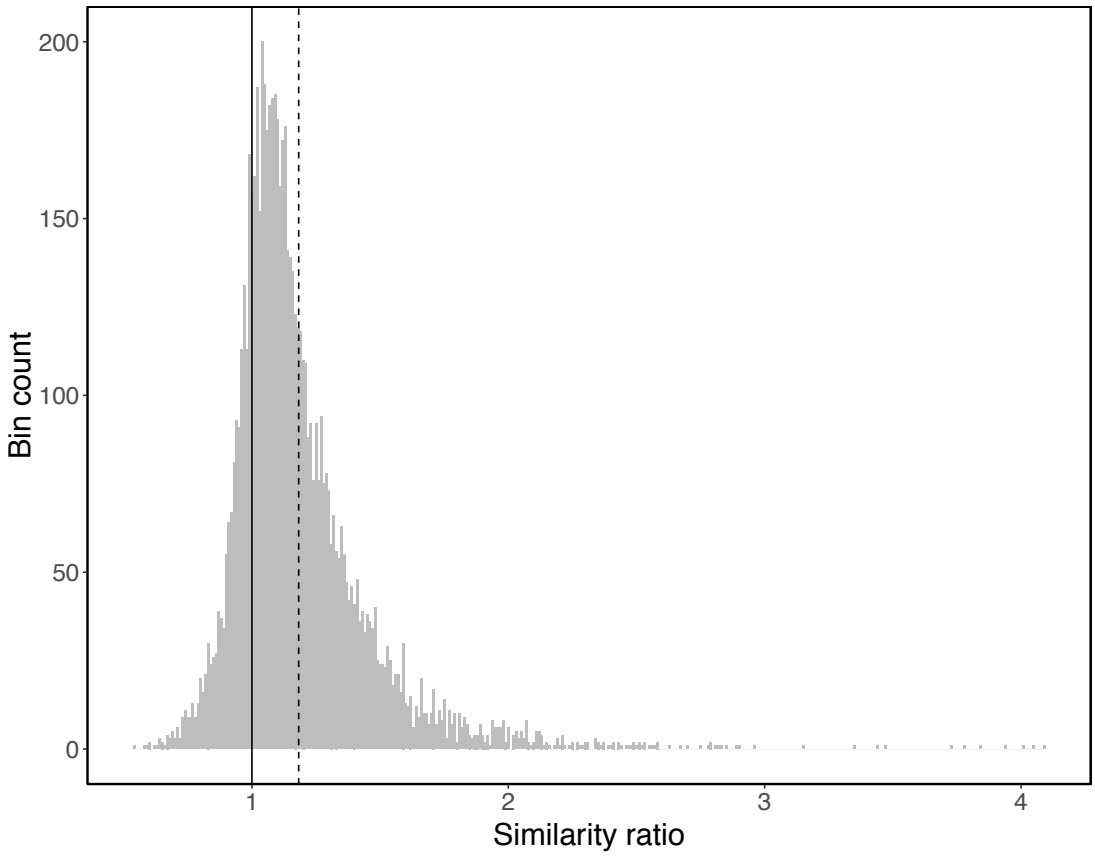

Figure S20. In *A. thaliana*, the phenotypes in the mutational neighborhoods of neutral neighbors are more similar than those of neutral pairs that are not neighbors. The distribution of the similarity ratio (Eq. 9) of the phenotype probability distributions (Eq. 6) is shown for neutral neighbors (n_1 and n_2) and for neutral pairs that are not neighbors (n_1 and n_3). For this analysis, we considered all 7,098 pairs of neutral neighbors in the genotype network for AZF2, and sampled the same number of neutral pairs that are not neighbors. The mutational neighborhoods of n_1 and n_2 are more similar than those of n_1 and n_3, because the mean of the ratio (vertical dashed line) is larger than unity (vertical solid line). The standard error of this mean (0.003) is minute compared to the difference between the mean and unity (0.182).

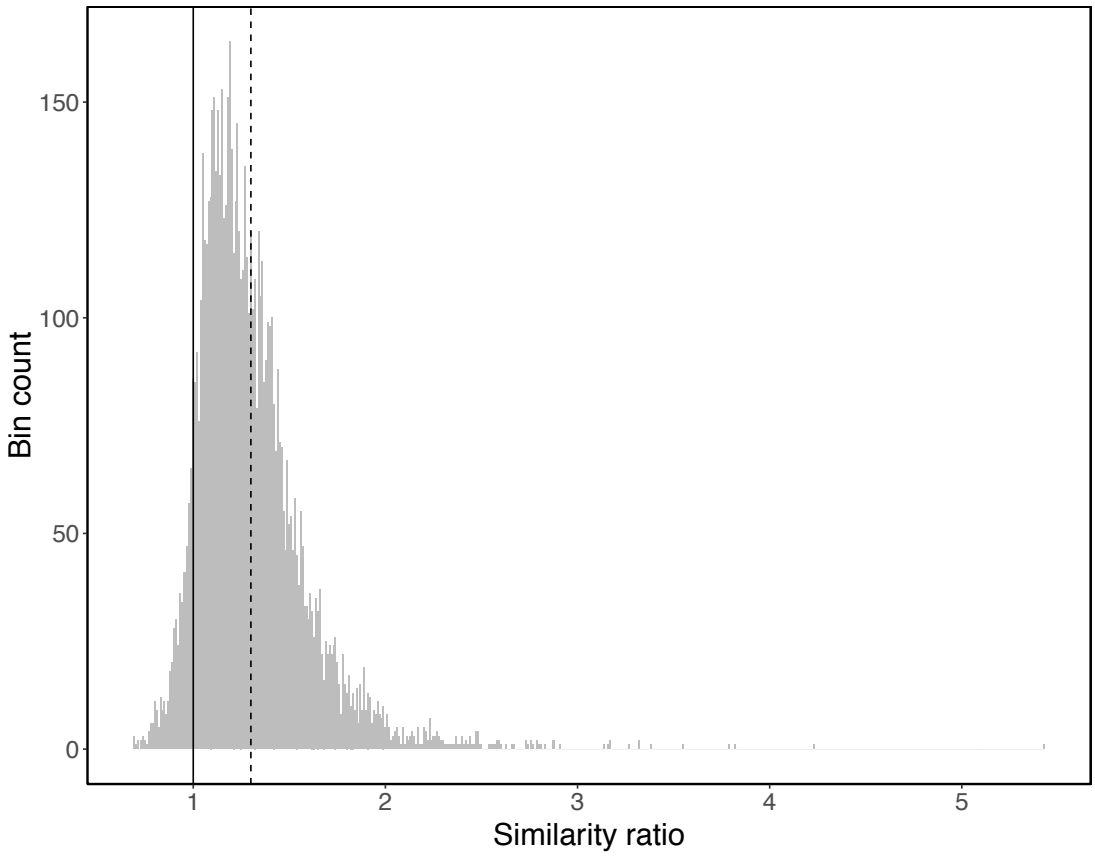

Figure S21. In *N. crassa*, the phenotypes in the mutational neighborhoods of neutral neighbors are more similar than those of neutral pairs that are not neighbors. The distribution of the similarity ratio (Eq. 9) of the phenotype probability distributions (Eq. 6) is shown for neutral neighbors (n_1 and n_2) and for neutral pairs that are not neighbors (n_1 and n_3). For this analysis, we considered all 7,379 pairs of neutral neighbors in the genotype network for NCU02525, and sampled the same number of neutral pairs that are not neighbors. The mutational neighborhoods of n_1 and n_2 are more similar than those of n_1 and n_3, because the mean of the ratio (vertical dashed line) is larger than unity (vertical solid line). The standard error of this mean (0.003) is minute compared to the difference between the mean and unity (0.3).

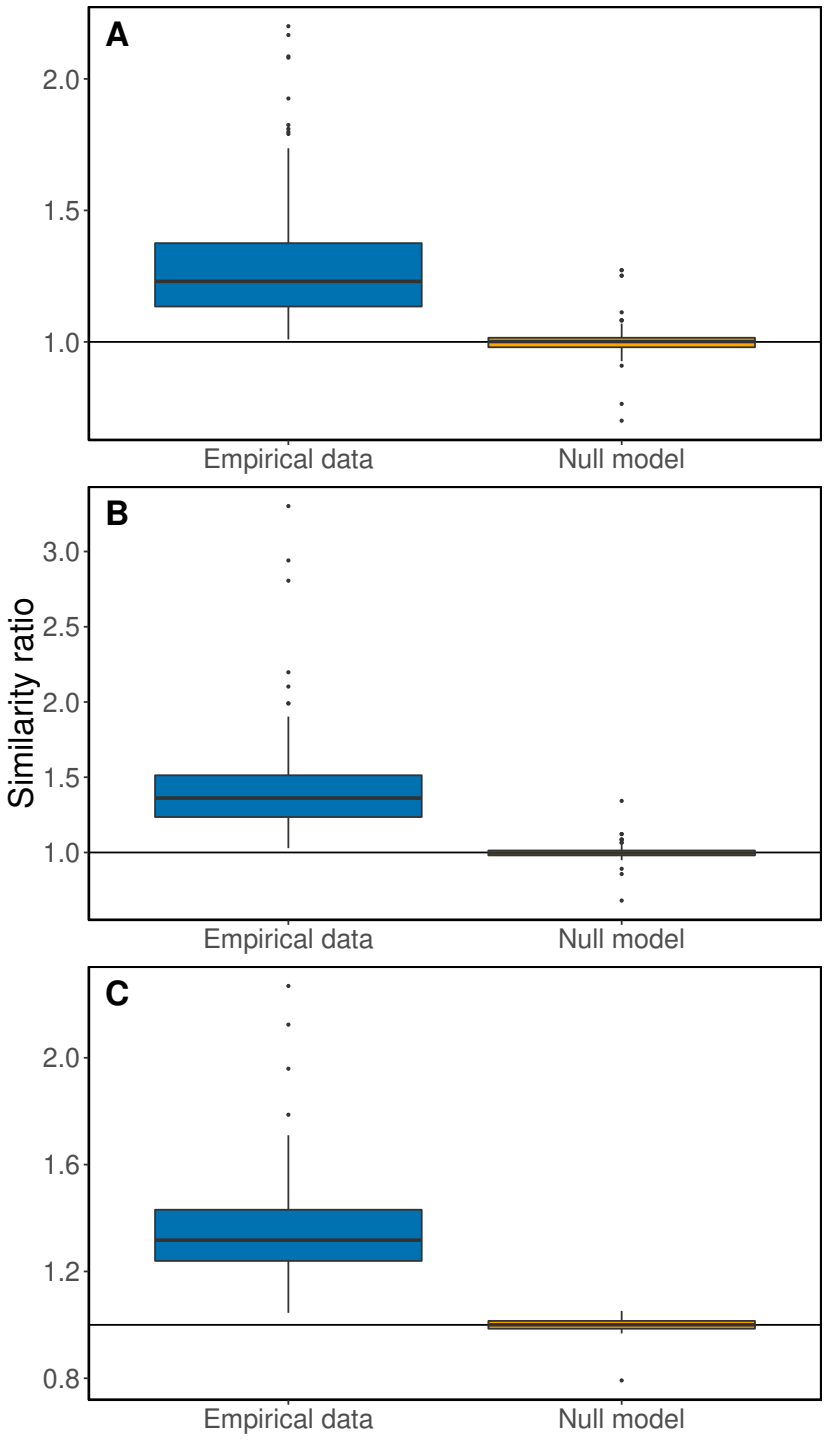

Figure S22. Comparison of similarity ratios in the empirical data to those of a null model. The phenotypes in the mutational neighborhoods of neutral neighbors are more similar than those of neutral pairs that are not neighbors: The median similarity ratio is larger than unity (horizontal solid line). This does not happen with the null model in any of the three species: (A) *A. thaliana*, (B) *M. musculus*, or (C) *N. crassa*. The empirical data is the same as presented in Figs. S19A, S20A, and S21A.

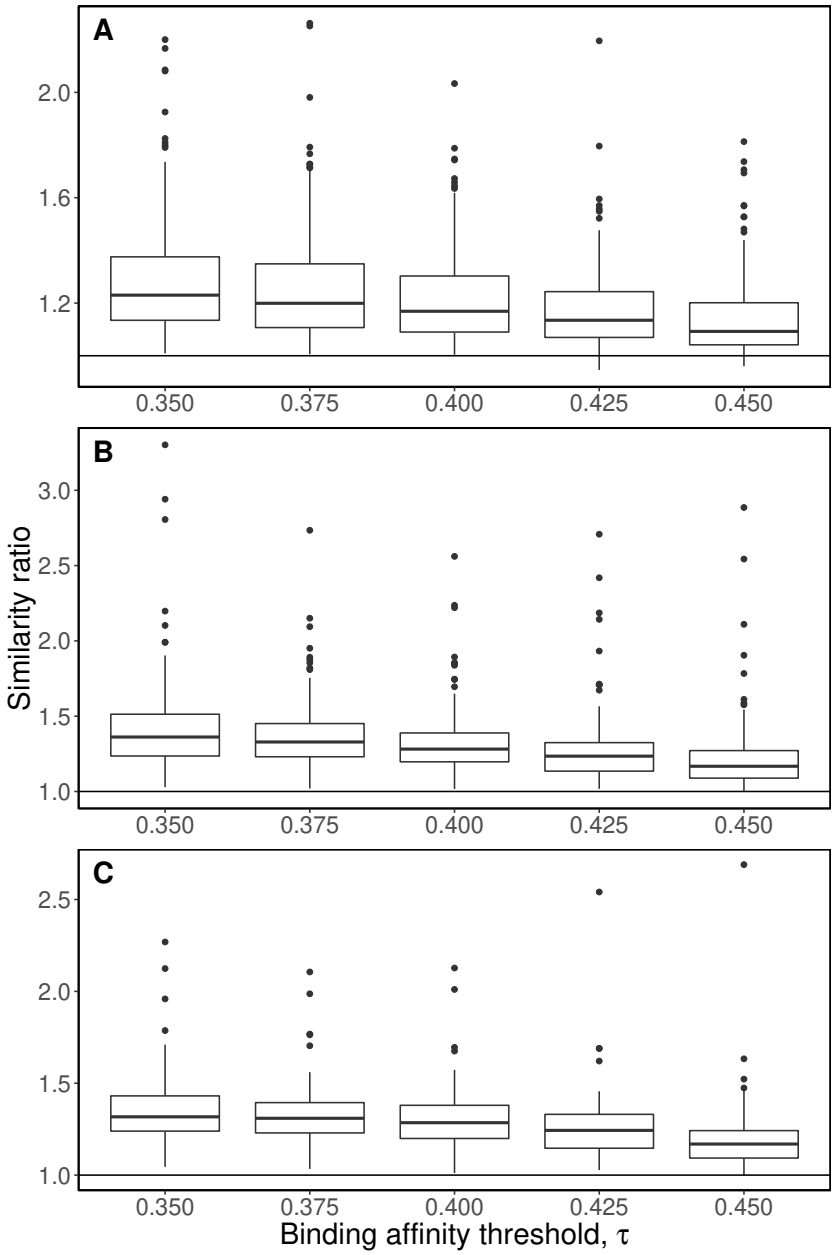

Figure S23. Similarity ratios with different binding affinity thresholds. The phenotypes in the mutational neighborhoods of neutral neighbors are more similar than those of neutral pairs that are not neighbors: The median similarity ratio is larger than unity (horizontal solid line) for all binding affinity thresholds. This observation does not change with binding affinity threshold in (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*. These data represent a sensitivity analysis of the results presented in Figs. S19A, S20A, and S21A.

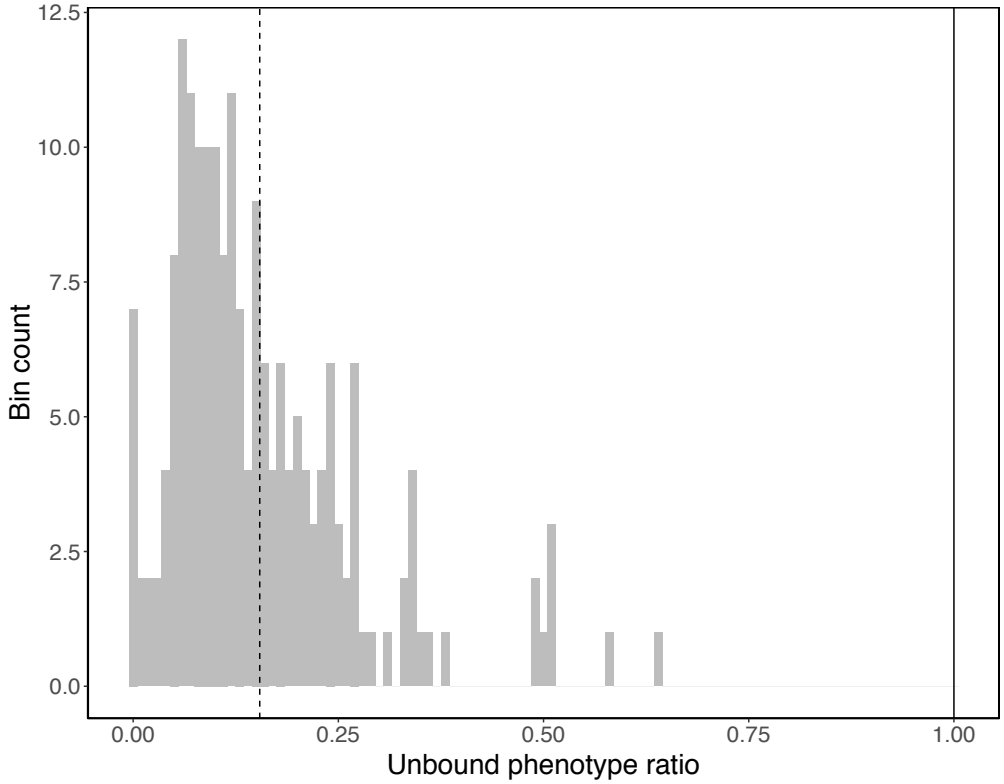

Figure S24. In *M. musculus*, unbound sites are underrepresented in the neighborhoods of bound sites. The distribution of the ratio $\phi_{\text{unbound},p}/f_{\text{unbound}}$, which is the probability of mutating from a sequence bound by TF p to an unbound sequence, divided by the null expectation of the frequency of unbound DNA sequences. The distribution is clearly skewed to values smaller than one, as shown by the distance of the distribution (vertical dashed line) to unity (vertical solid line).

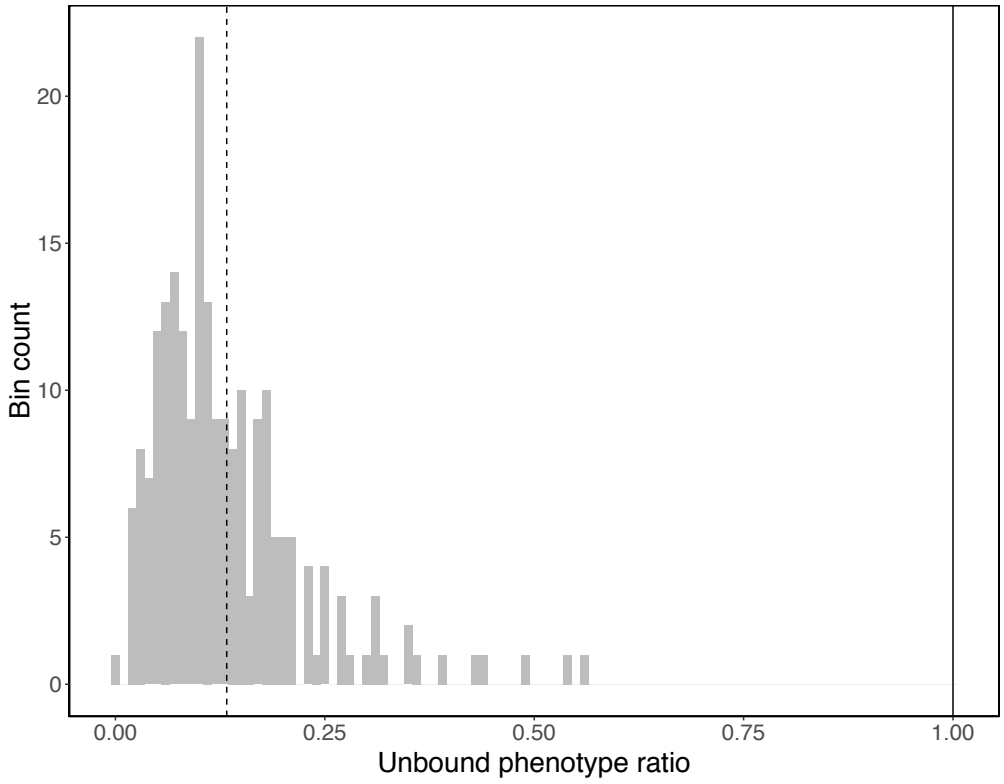

Figure S25. In *A. thaliana*, unbound sites are underrepresented in the neighborhoods of bound sites. The distribution of the ratio $\phi_{\text{unbound},p}/f_{\text{unbound}}$, which is the probability of mutating from a sequence bound by TF p to an unbound sequence, divided by the null expectation of the frequency of unbound DNA sequences. The distribution is clearly skewed to values smaller than one, as shown by the distance of the distribution (vertical dashed line) to unity (vertical solid line).

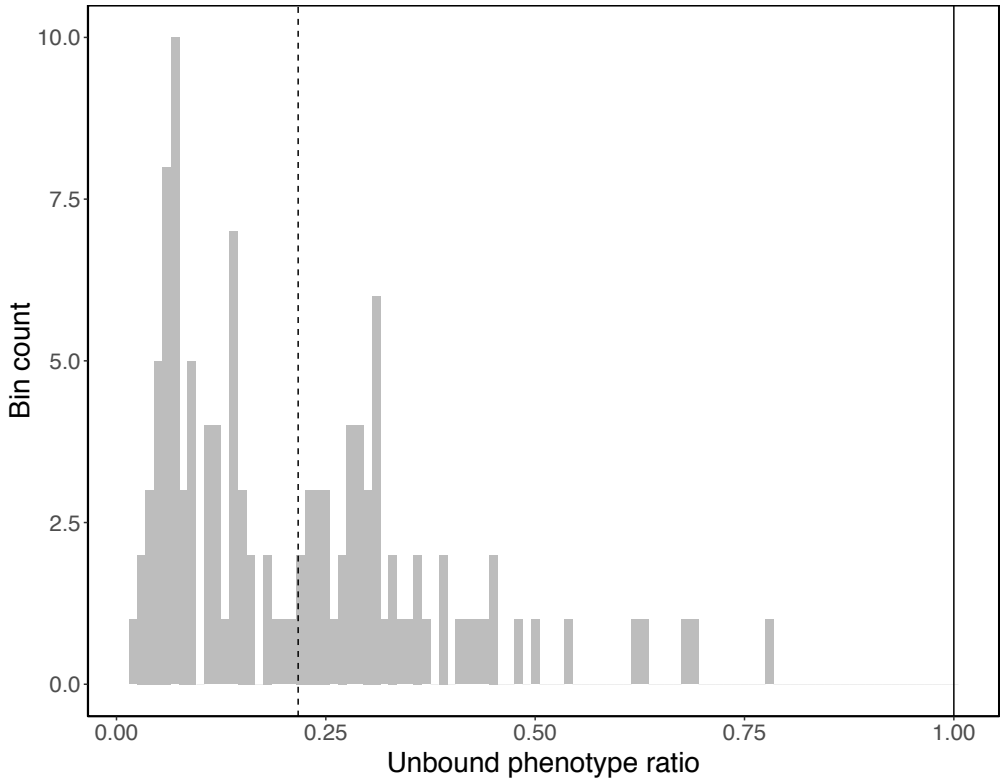

Figure S26. In *N. crassa*, unbound sites are underrepresented in the neighborhoods of bound sites. The distribution of the ratio $\phi_{\text{unbound},p}/f_{\text{unbound}}$, which is the probability of mutating from a sequence bound by TF p to an unbound sequence, divided by the null expectation of the frequency of unbound DNA sequences. The distribution is clearly skewed to values smaller than one, as shown by the distance of the distribution (vertical dashed line) to unity (vertical solid line).

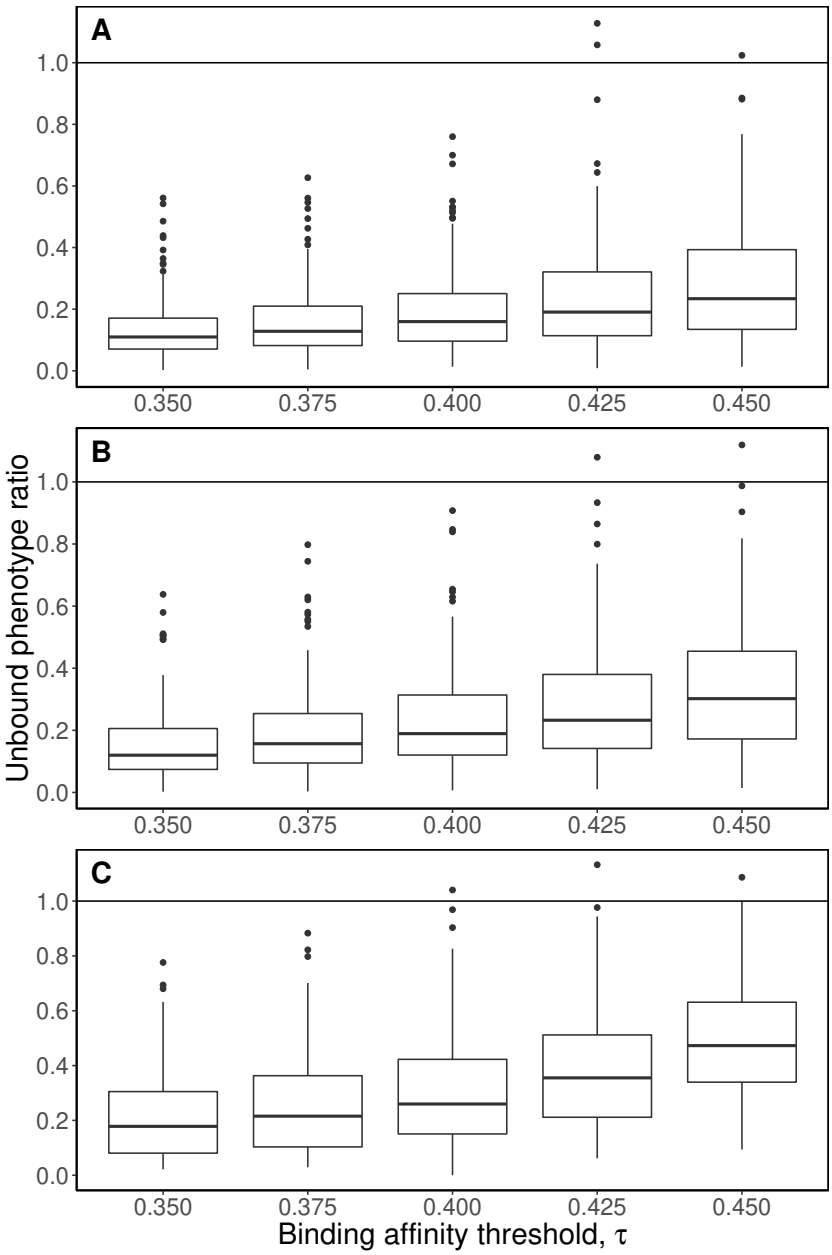

Figure S27. Unbound sites are underrepresented in the neighborhoods of bound sites for all binding affinity thresholds. The distribution of the ratio $\phi_{\text{unbound},p}/f_{\text{unbound}}$, which is the probability of mutating from a sequence bound by TF p to an unbound sequence, divided by the null expectation of the frequency of unbound DNA sequences. The distribution is skewed toward values smaller than one, as shown by the distance of the median of the distribution to unity (horizontal solid line). This observation does not change with binding affinity threshold in (A) *A. thaliana*, (B) *M. musculus*, and (C) *N. crassa*. These data represent a sensitivity analysis of the results presented in Figs. S24A, S25A, and S26A.

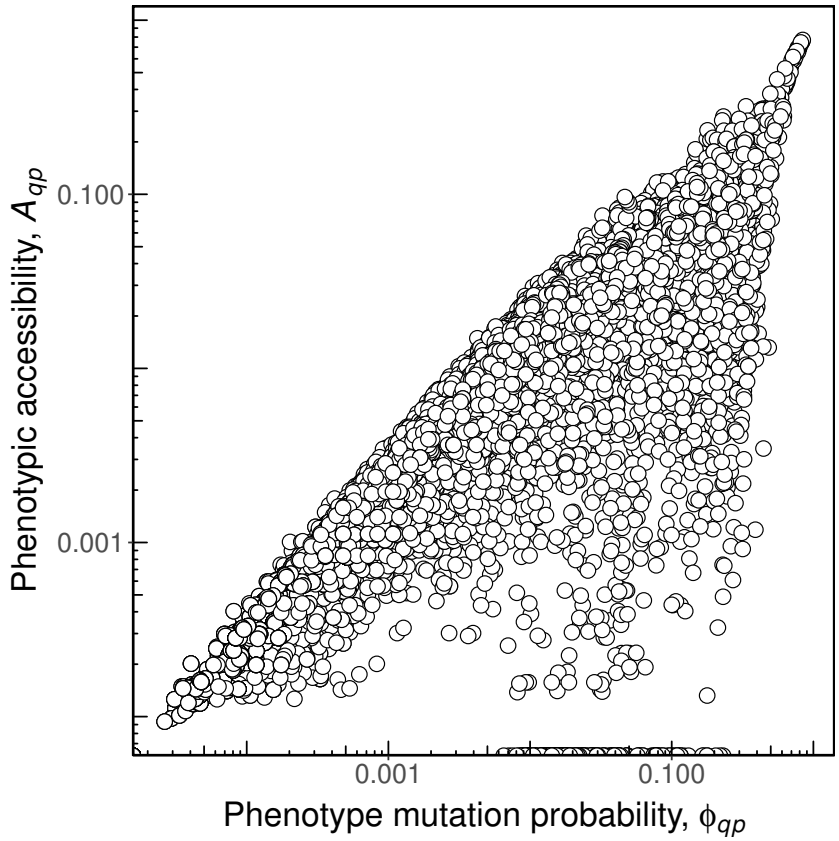

Figure S28. Phenotypic accessibility A_{qp} is strongly correlated with ϕ_{qp} (Spearman's $r = 0.95$, $p < 10^{-6}$). Each circle represents one of the 35,910 pairs of TFs from *M. musculus*. Half circles at the bottom of the panel denote pairs of phenotypes with phenotypic accessibility = 0. Note the logarithmic scale on both axes.

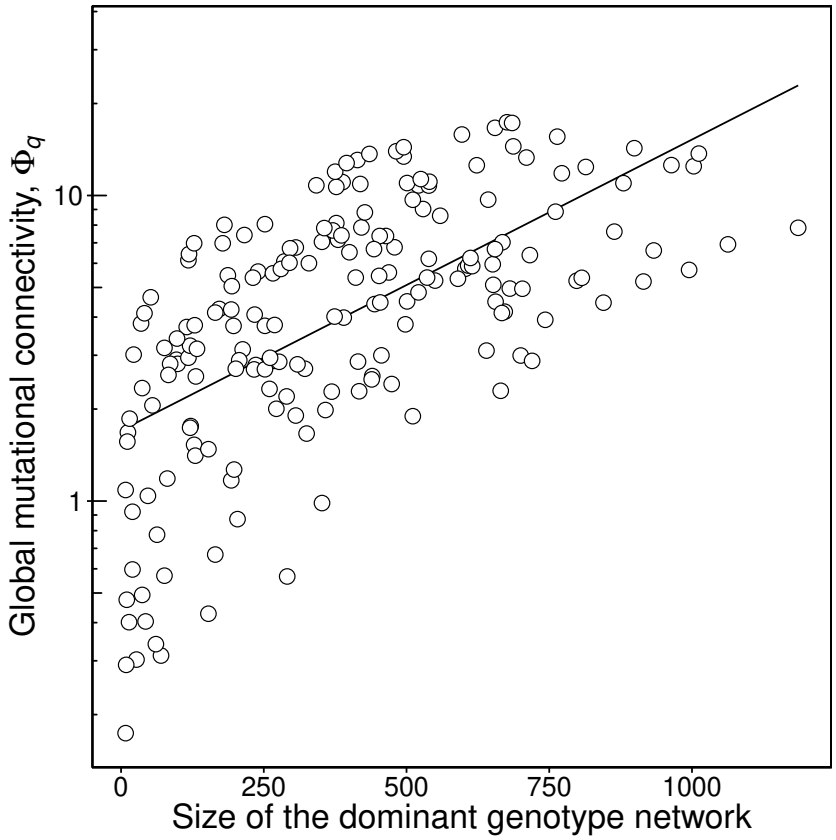

Figure S29. In *M. musculus*, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network. Each circle shows the global mutational connectivity Φ_q of one of the 190 *M. musculus* TFs, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual guide.

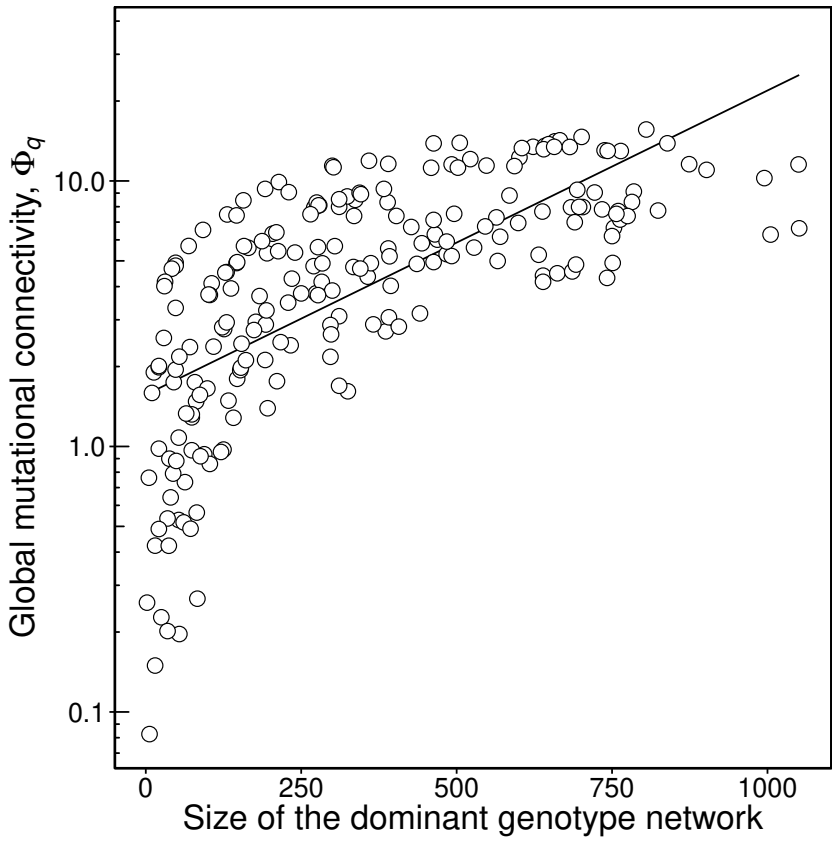

Figure S30. In *A. thaliana*, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network. Each circle shows the global mutational connectivity Φ_q of one of the 217 *A. thaliana* TFs, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual guide.

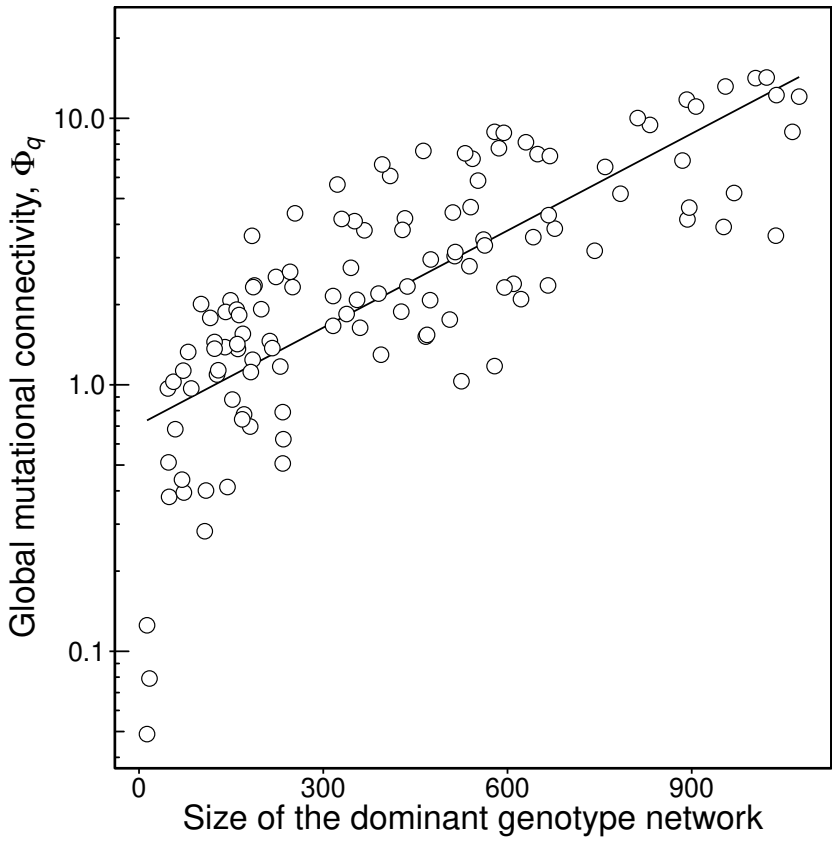

Figure S31. In *N. crassa*, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network. Each circle shows the global mutational connectivity Φ_q of one of the 118 *N. crassa* TFs, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual aid.

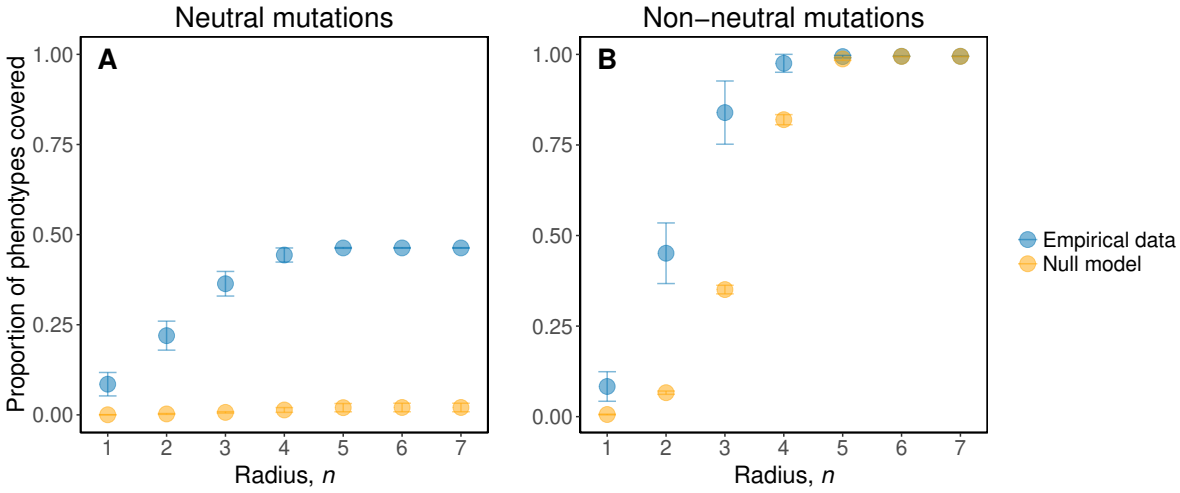

Figure S32. Comparison of phenotype space covering in the empirical data to that in the null model. The proportion of phenotypes covered as a function of the mutational radius n from a given binding site, averaged across all binding sites of the murine TF Sp110. The comparison between empirical data and the null model is made for both (A) neutral mutations and (B) non-neutral mutations. Error bars are the standard deviations of the mean. The empirical data is the same as presented in Fig. 4A.

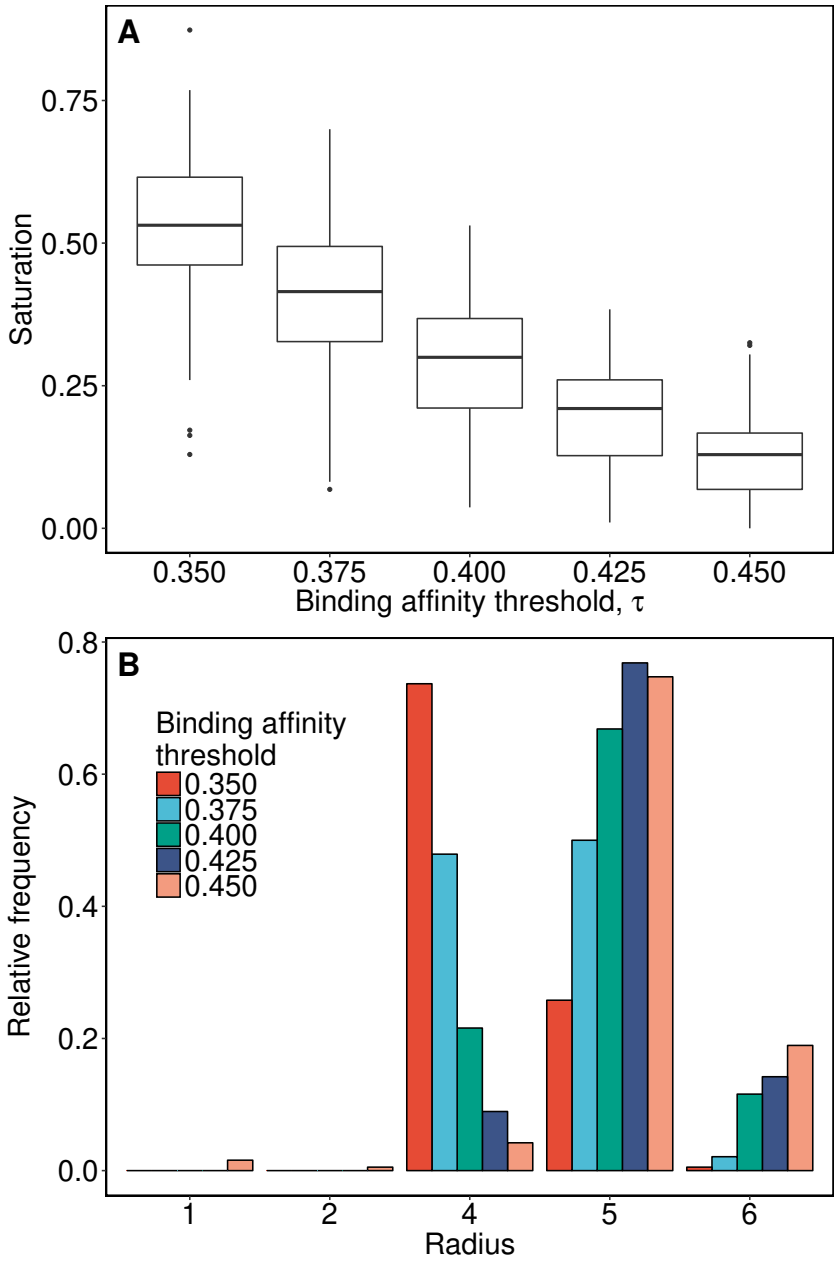

Figure S33. Phenotype space covering with different binding affinity thresholds for 190 TFs from *M. musculus*. (A) The maximum proportion of phenotypes covered by neutral mutations as a function of the binding affinity threshold, for all 190 murine TFs. (B) The distribution of the average mutational radius n that covers all phenotypes in genotype space for different binding affinity thresholds.

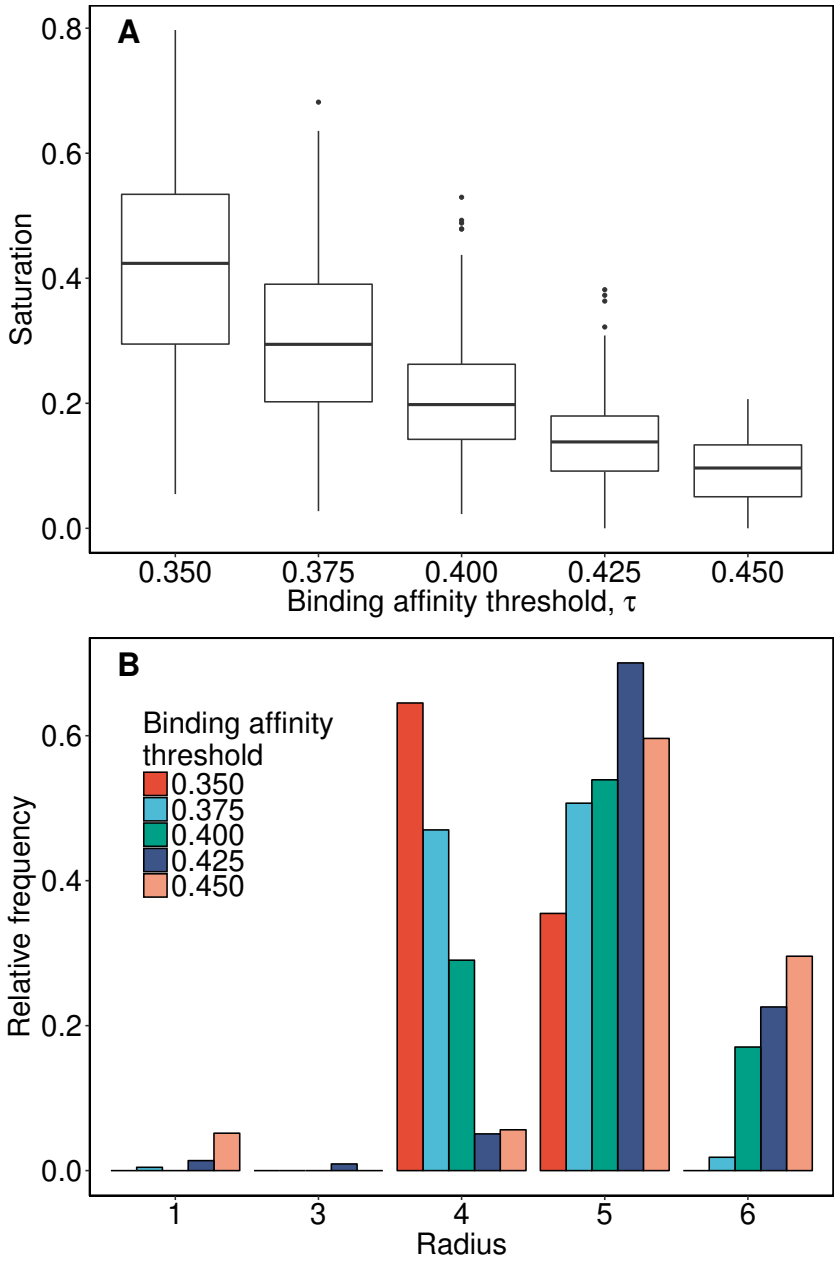

Figure S34. Phenotype space covering with different binding affinity thresholds for 217 TFs from *A. thaliana*. (A) The maximum proportion of phenotypes covered by neutral mutations as a function of the binding affinity threshold, for all 217 plant TFs. (B) The distribution of the average mutational radius n that covers all phenotypes in genotype space for different binding affinity thresholds.

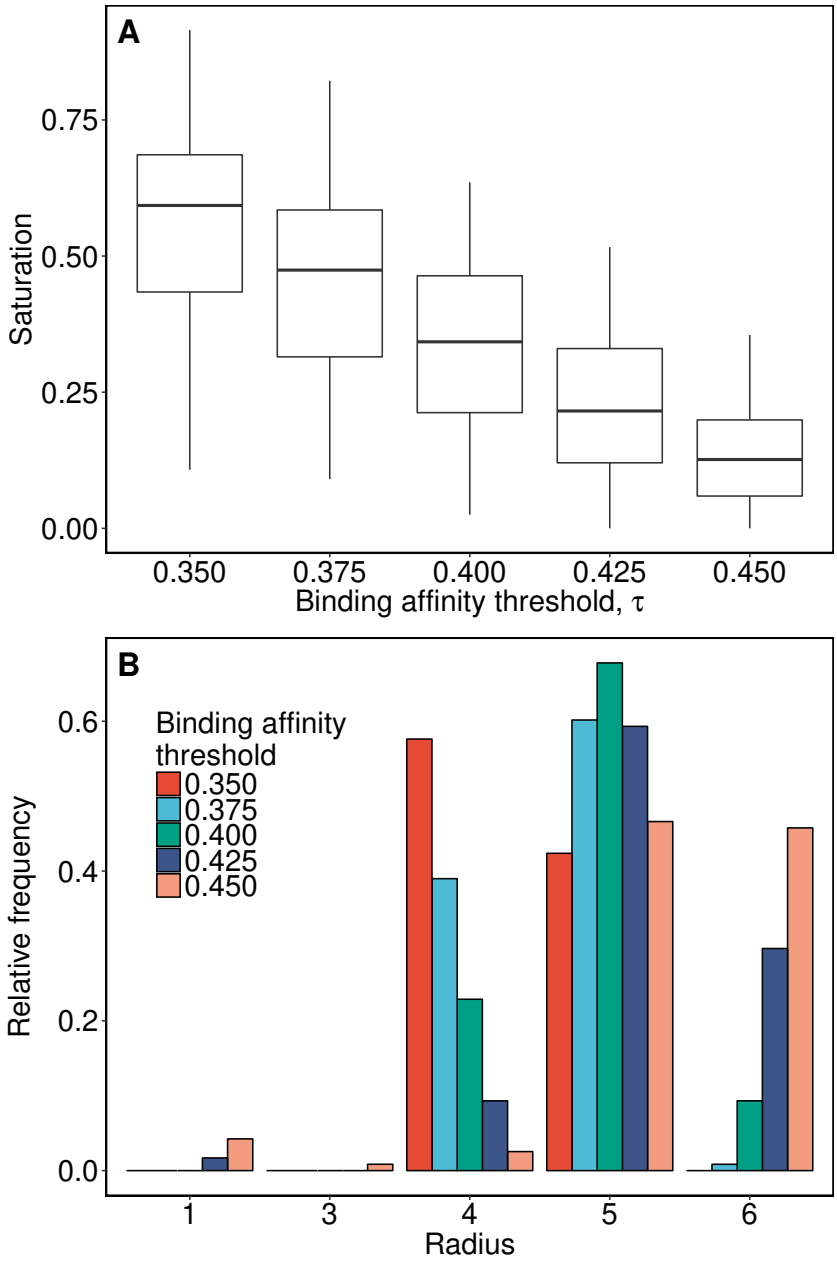

Figure S35. Phenotype space covering with different binding affinity thresholds for 118 TFs from *N. crassa*. (A) The maximum proportion of phenotypes covered by neutral mutations as a function of the binding affinity threshold, for all 118 fungal TFs. (B) The distribution of the average mutational radius n that covers all phenotypes in genotype space for different binding affinity thresholds.

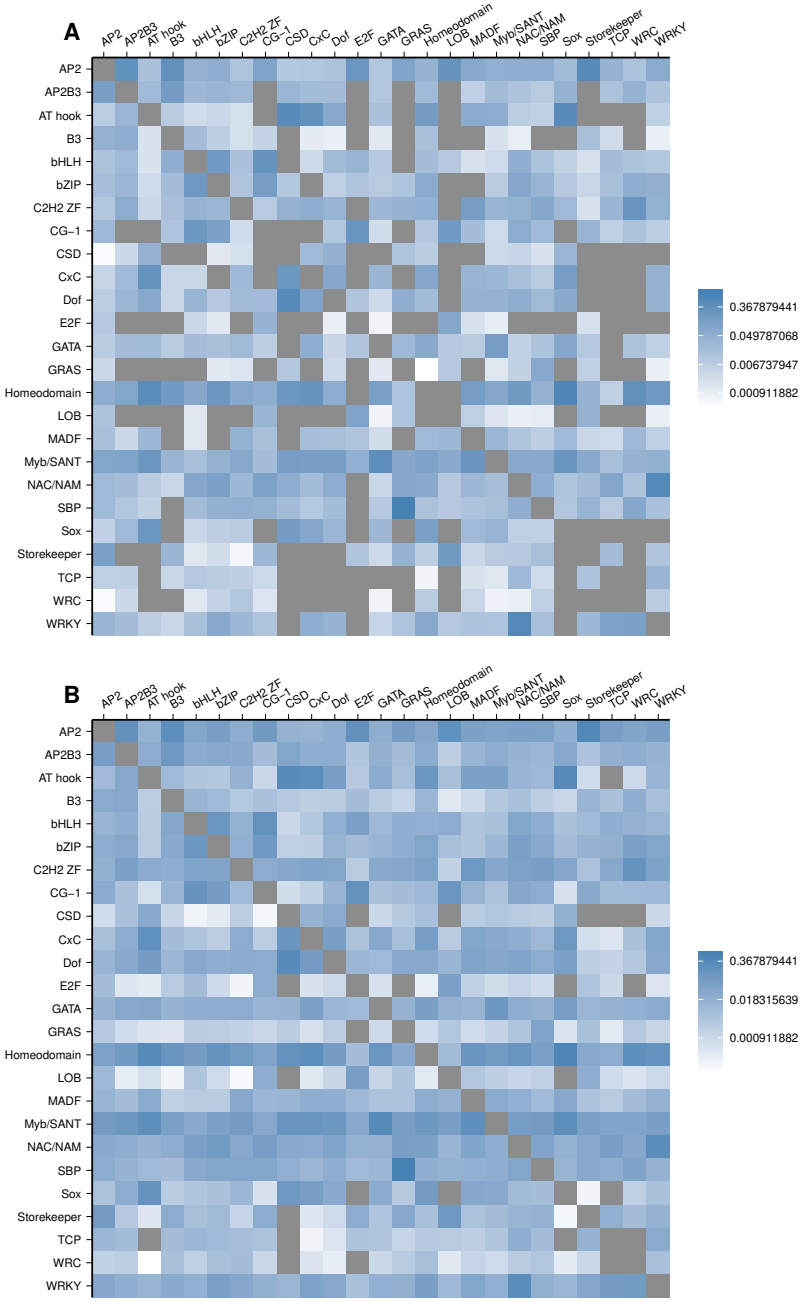

Figure S36. Matrices of internetwork relationships for the genotype networks of DNA-binding domains from *A. thaliana*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. Each row and column represents a different DNA-binding domain genotype network. Domains are ordered alphabetically. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

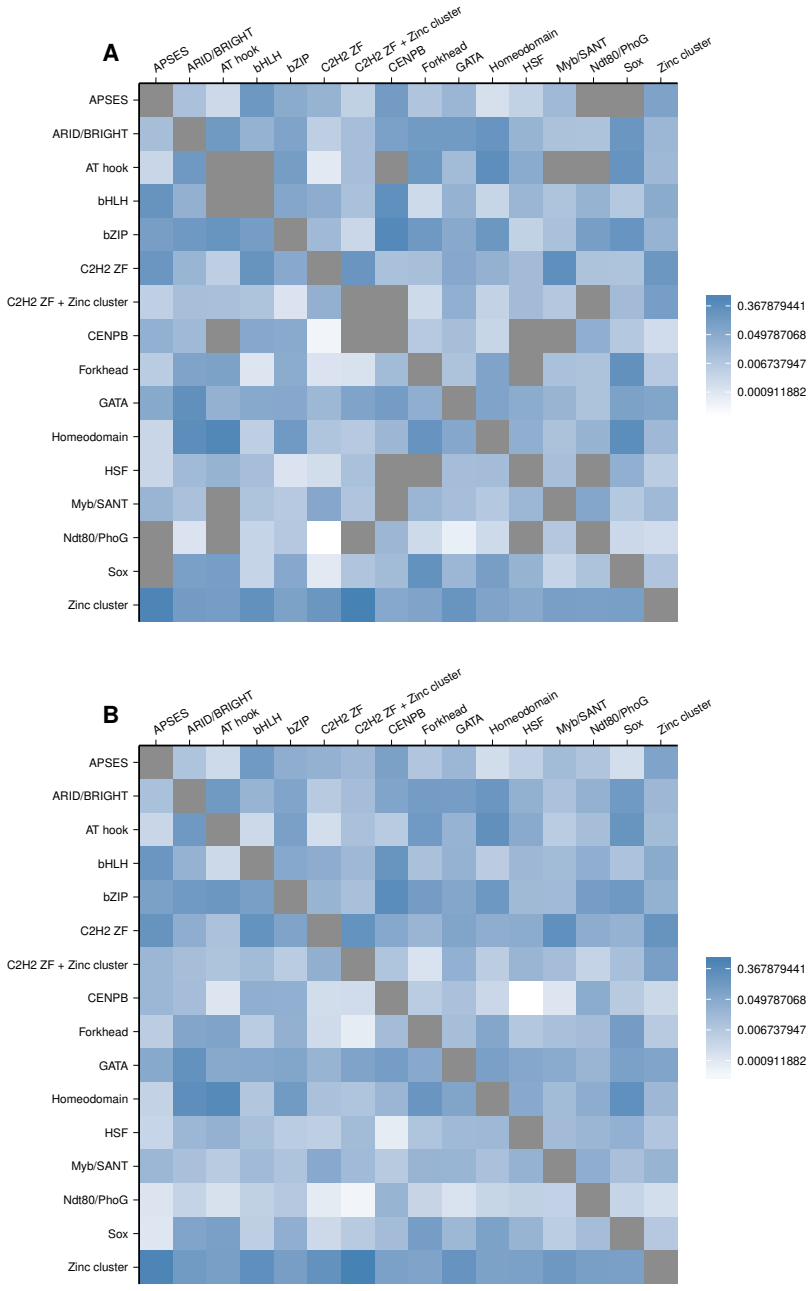

Figure S37. Matrices of internetwork relationships for the genotype networks of DNA-binding domains from *N. crassa*. Heatmaps of log10-transformed (A) overlap and (B) ϕ_{qp}, the probability of mutating from the genotype network of phenotype p to the genotype network of phenotype q. Each row and column represents a different DNA-binding domain genotype network. Domains are ordered alphabetically. Cells colored in gray indicate either N/A values (on the diagonal) or values equal to zero (off-diagonal).

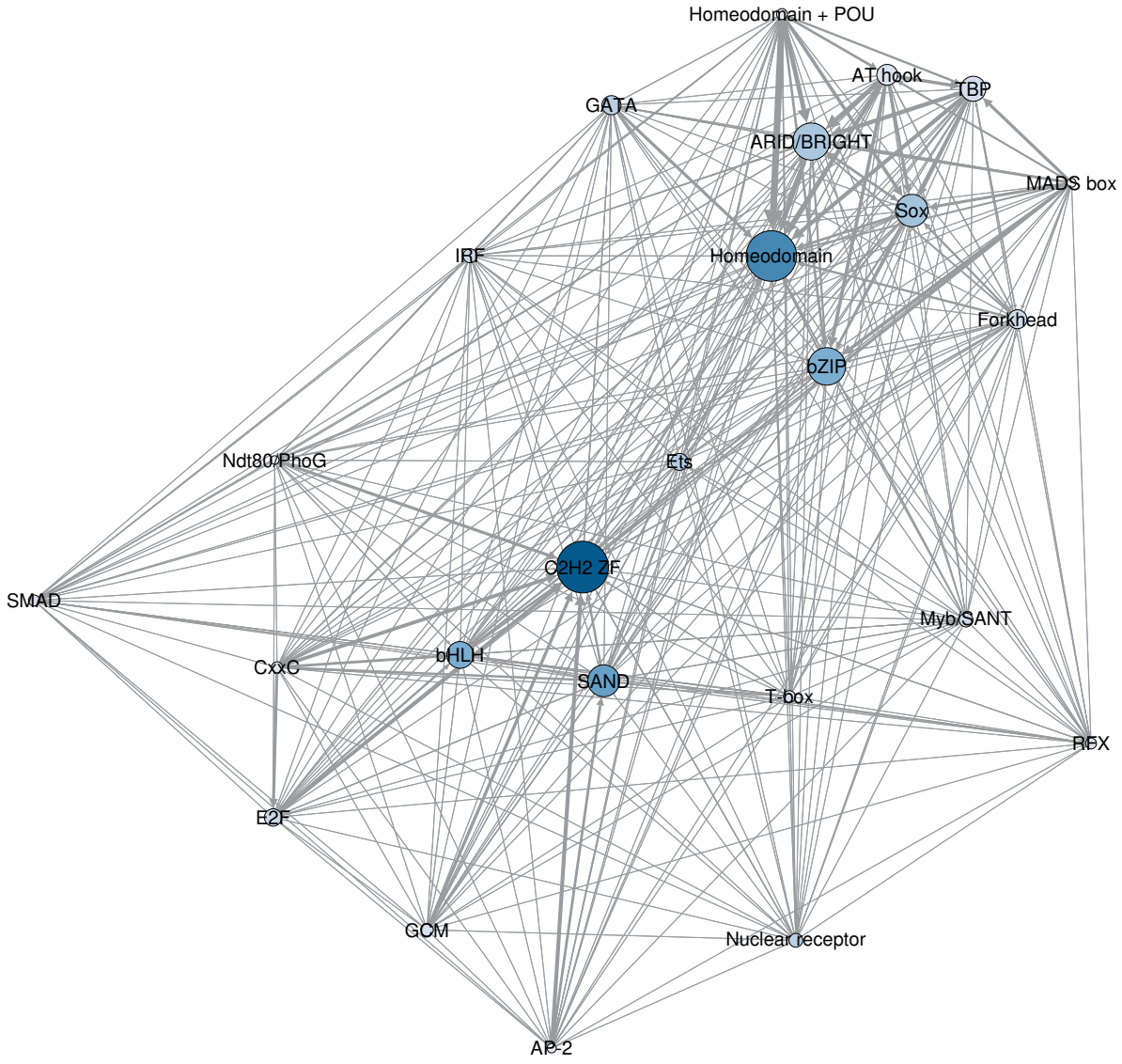

Figure S38. Phenotype network for 25 DNA-binding domains from *M. musculus*. The nodes in this network represent the dominant genotype networks of DNA binding domains, and edges connect nodes if their corresponding genotype networks are connected by at least one non-neutral mutation. The size of the edges is proportional to the ϕ_{qp} among domains. Node size is proportional to the size of the associated genotype network. Node color represents the global mutational connectivity Φ_q of each domain (darker nodes have larger Φ_q).

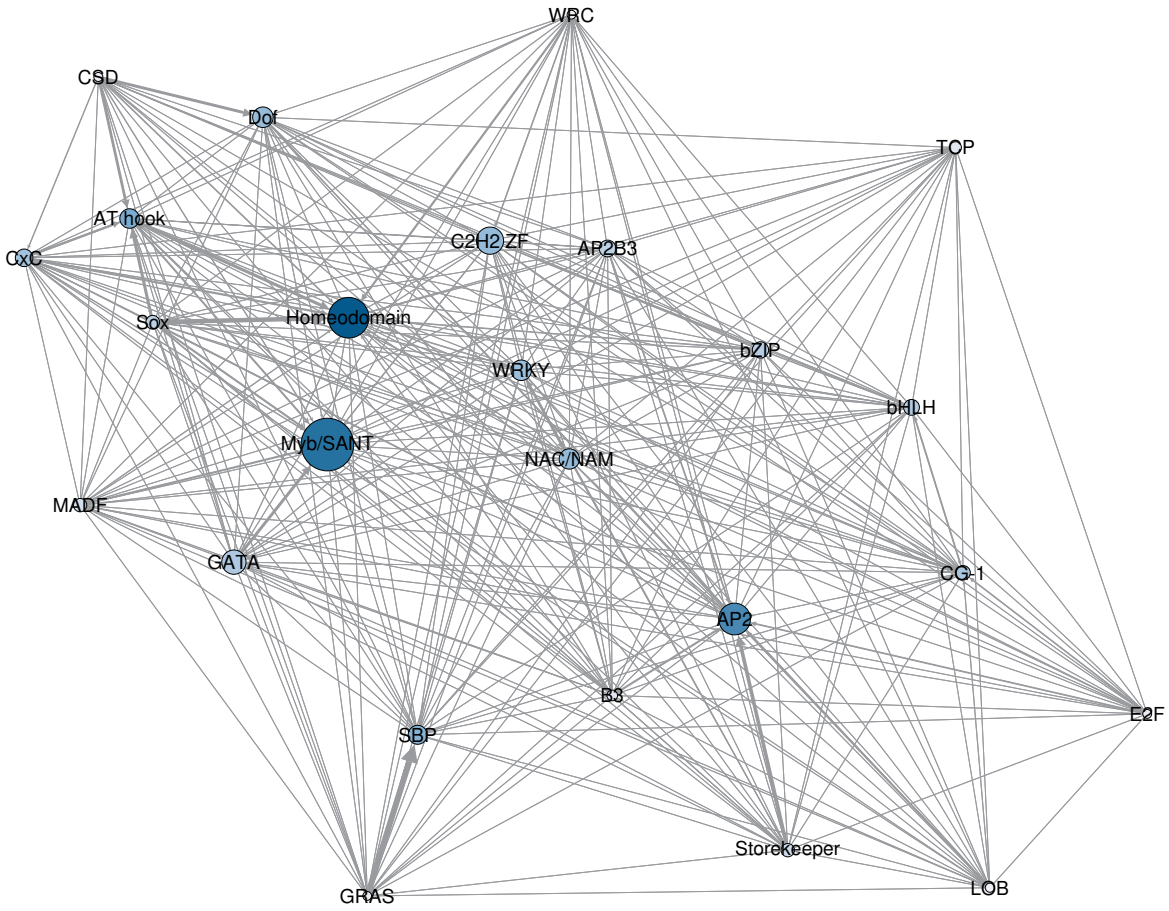

Figure S39. Phenotype network for 25 DNA-binding domains from *A. thaliana*. The nodes in this network represent the dominant genotype networks of DNA binding domains, and edges connect nodes if their corresponding genotype networks are connected by at least one non-neutral mutation. The size of the edges is proportional to the ϕ_{qp} among domains. Node size is proportional to the size of the associated genotype network. Node color represents the global mutational connectivity Φ_q of each domain (darker nodes have larger Φ_q).

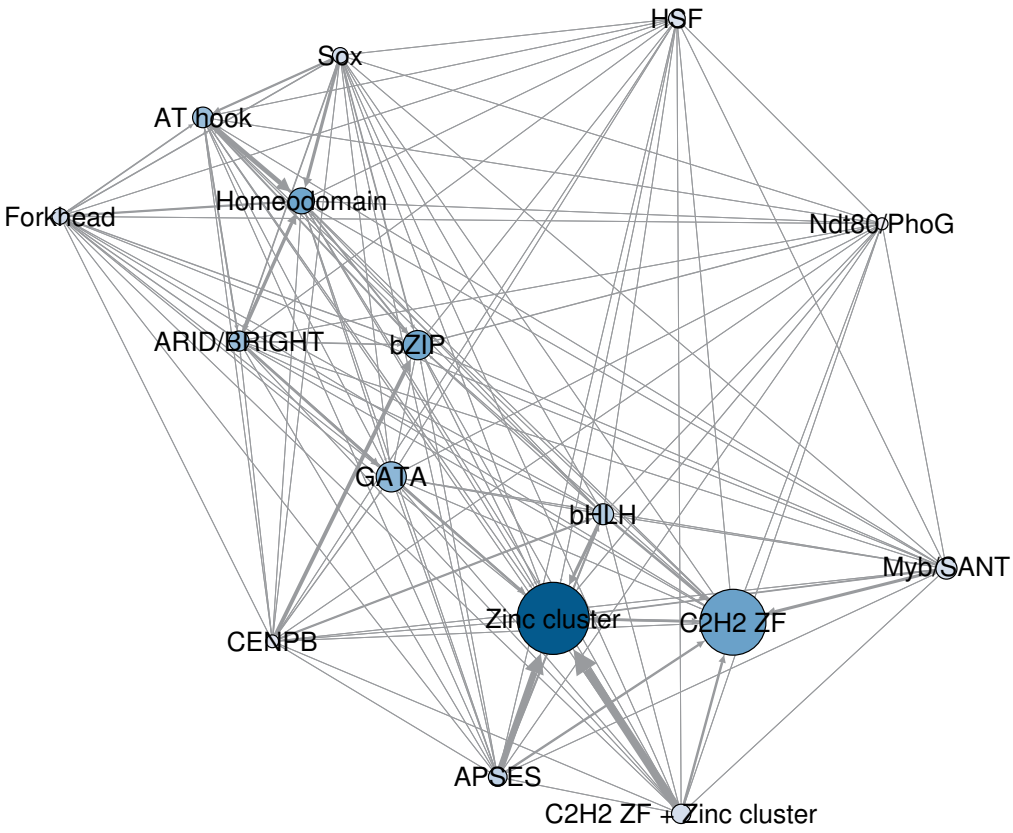

Figure S40. Phenotype network for 16 DNA-binding domains from *N. crassa*. The nodes in this network represent the dominant genotype networks of DNA binding domains, and edges connect nodes if their corresponding genotype networks are connected by at least one non-neutral mutation. The size of the edges is proportional to the ϕ_{qp} among domains. Node size is proportional to the size of the associated genotype network. Node color represents the global mutational connectivity Φ_q of each domain (darker nodes have larger Φ_q).

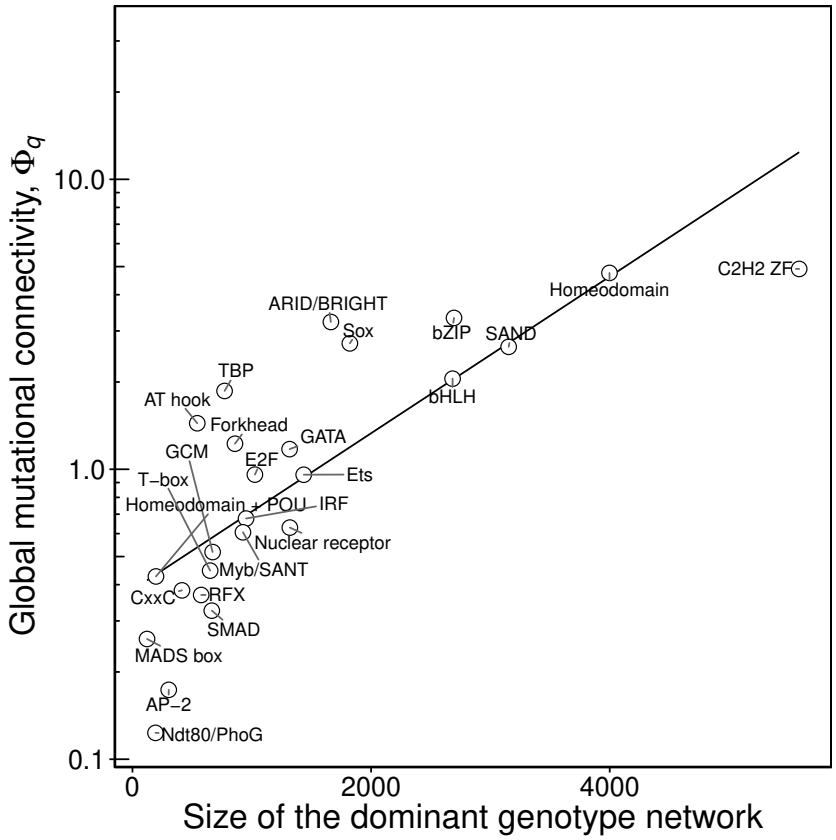

Figure S41. In *M. musculus*, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network. Each circle shows the global mutational connectivity Φ_q of one of the 25 *M. musculus* DNA binding domains, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual aid.

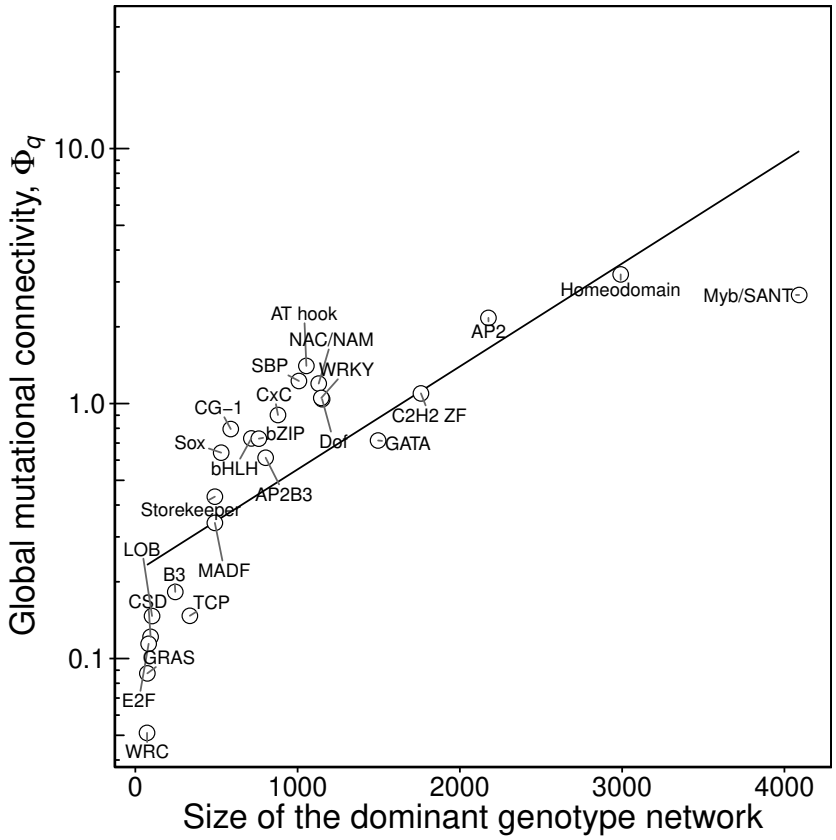

Figure S42. In *A. thaliana***, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network.** Each circle shows the global mutational connectivity Φ_q of one of the 25 *A. thaliana* DNA binding domains, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual aid.

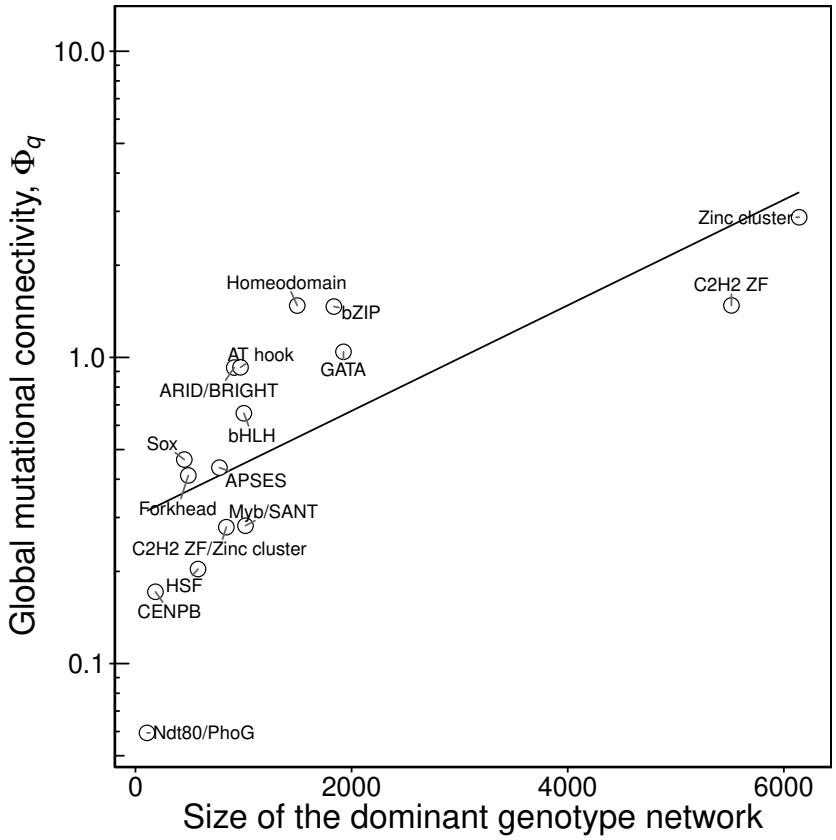

Figure S43. In *N. crassa*, the global mutational connectivity of a phenotype increases with the size of its dominant genotype network. Each circle shows the global mutational connectivity Φ_q of one of the 16 *N. crassa* DNA binding domains, as a function of the number of binding sites in its dominant genotype network. The solid line is the best linear fit to the data and is provided as a visual aid.

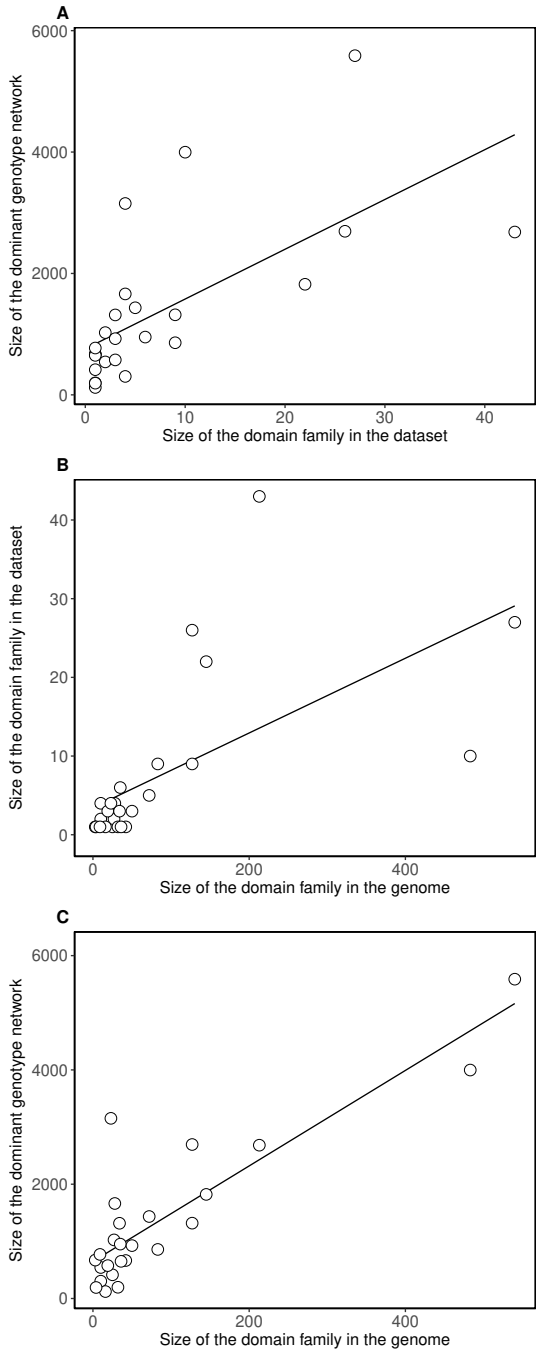

Figure S44. Binding domains with more TFs have larger genotype networks in *M. musculus***.** (A) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per domain in our dataset (Spearman's $r = 0.8$, $p = 2 \times 10^{-6}$). (B) The relationship between the number of TFs per binding domain in our dataset and the number of TFs per binding domain in the *M. musculus* genome (Spearman's $r = 0.75$, $p = 1.4 \times 10^{-5}$). (C) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per binding domain in the *M. musculus* genome (Spearman's $r = 0.7$, $p = 9.6 \times 10^{-5}$). In each panel, each circle represents one of the 25 *M. musculus* binding domains in our dataset.

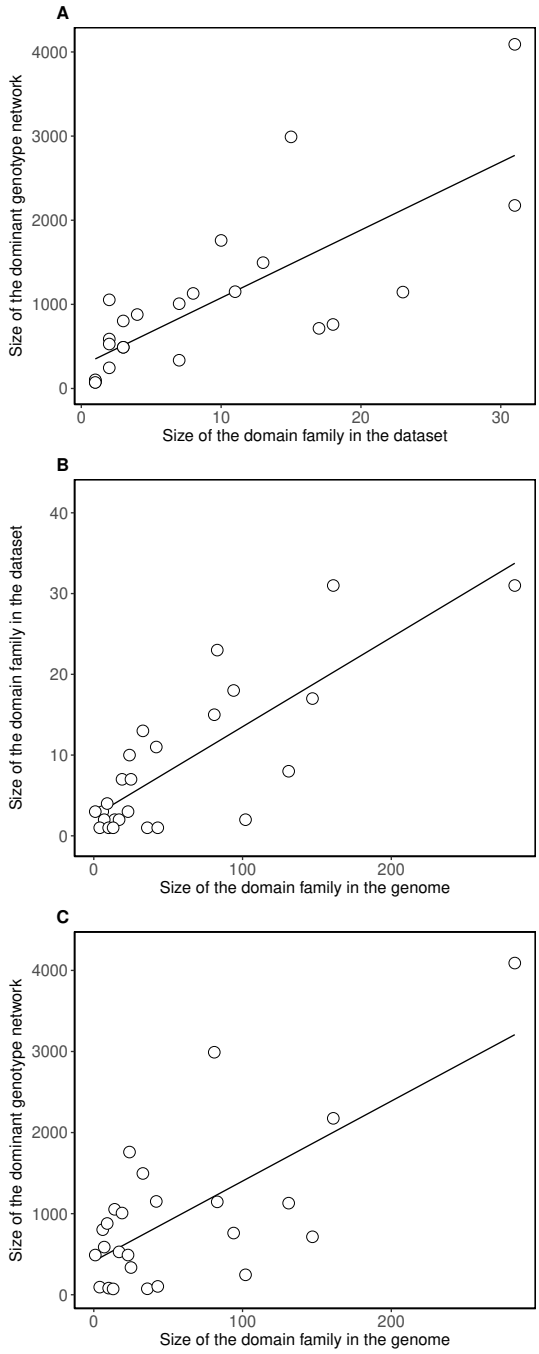

Figure S45. Binding domains with more TFs have larger genotype networks in *A. thaliana*. (A) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per domain in our dataset (Spearman's $r = 0.83$, $p = 2.8 \times 10^{-7}$). (B) The relationship between the number of TFs per binding domain in our dataset and the number of TFs per binding domain in the *A. thaliana* genome (Spearman's $r = 0.64$, $p = 5.8 \times 10^{-4}$). (C) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per binding domain in the *A. thaliana* genome (Spearman's $r = 0.44$, $p = 2.9 \times 10^{-2}$). In each panel, each circle represents one of the 25 *A. thaliana* binding domains in our dataset.

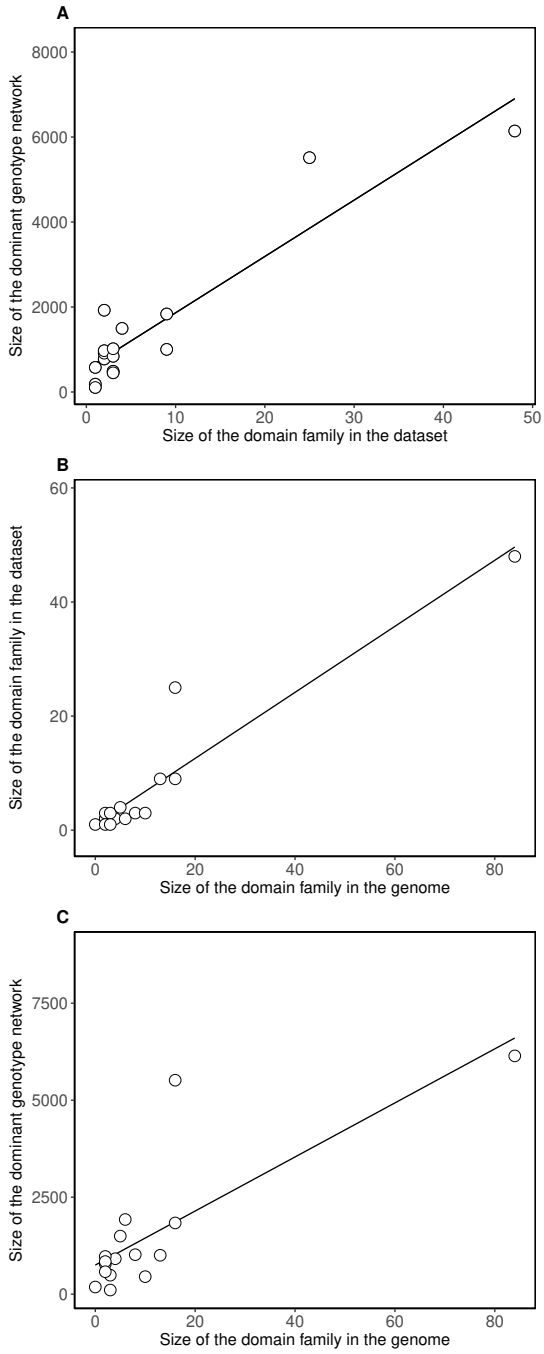

Figure S46. Binding domains with more TFs have larger genotype networks in *N. crassa*. (A) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per domain in our dataset (Spearman's $r = 0.93$, $p = 2 \times 10^{-7}$). (B) The relationship between the number of TFs per binding domain in our dataset and the number of TFs per binding domain in the *N. crassa* genome (Spearman's $r = 0.94$, $p = 4.8 \times 10^{-8}$). (C) The relationship between the size of a binding domain's dominant genotype network and the number of TFs per binding domain in the *N. crassa* genome (Spearman's $r = 0.8$, $p = 2.3 \times 10^{-4}$). In each panel, each circle represents one of the 16 *N. crassa* binding domains in our dataset.